University of Alberta

New Methods for the Characterization of Pharmaceutically Important Proteins

by

Yunjun Wang ©

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Faculty of
Pharmacy and Pharmaceutical of Sciences
Edmonton, Alberta
FALL, 1999

National Library
of Canada

Acquisitions and
Bibliographic Services

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale
du Canada

Acquisitions et
services bibliographiques

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-46942-5

Canada

University of Alberta
Library Release Form

Name of Author: Yunjun Wang

Title of Thesis: New Methods for the Characterization of Pharmaceutically Important Proteins

Degree: Doctor of Philosophy

Year this Degree Granted: 1999

Permission is hereby granted to the University of Alberta Library to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly, or scientific research purposes only.

The author reserves all other publication and other rights in association with the copyright in the thesis, and except as hereinbefore provided, neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatever without the author's prior written permission.

Faculty of Pharmacy and
Pharmaceutical Sciences
University of Alberta
Edmonton, Alberta
T6G 2C2

Date: July 6, 1999

University of Alberta

Faculty of Graduate Studies and Research

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled "New Methods for the Characterization of Pharmaceutically Important Proteins" submitted by Yunjun Wang in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

David Wishart

Dr. David S. Wishart

Brian Sykes

Dr. Brian D. Sykes

Dr. Mavanur R. Suresh

Dr. George. Kotovych

Dr. Len I. Wiebe

Dr. Lawrence P. Mcintosh

Date: July 5 1999

Abstract

In chapters 2 and 3 a simple linear relationship between the $^3J_{HNH\alpha}$ coupling constant and the half-height linewidth ($\Delta\nu_{1/2}$) of in-phase NMR peaks is described. This relationship permits the rapid and accurate determination of $^3J_{HNH\alpha}$ from both homonuclear and heteronuclear NMR spectra. Testing on more than one thousand $^3J_{HNH\alpha}$ coupling constants from nearly twenty different peptide/protein NMR spectra shows that despite the method's simplicity, the results obtained are comparable in accuracy and precision to the best techniques published to date. In Chapter 4, solution structures of both reduced and oxidized T4 Grx, as determined by ^{1}H NMR techniques are described. In chapter 5, NMR structures of CbnB2, in 90% trifluoroethanol are reported. CbnB2 has a well-defined central helix (from residues 18 to 39), but a disordered N-terminus. Comparisons of the CbnB2 structure with the refined solution structure of Leucocin A (LeuA) – another type IIa bacteriocin – indicate that this helical structure is well conserved between the two peptides, but the N-terminal structure (i.e. the presumptive receptor binding site) is not. This result is of particular interest because LeuA and CbnB2 actually exhibit >66% sequence identity over their first 24 N-terminal residues. This suggests that the highly ordered N-terminal β-sheet structure seen in LeuA may either be an artifact of solution conditions or that the N-terminal structure in type IIa bacteriocins may not be critical to their function. Chapter 6 reports the high level production and efficient isotopic labeling of an anti-PSA single chain antibody (scFvB80) from *P. pastoris.* Competitive ELISAs show that purified scFvB80 has ~20% PSA binding activity relative to the full length antibody.Additional ^{1}H NMR studies show that scFvB80 also interacts with PSA epitope/mimetope peptides previously identified from studies of the whole antibody. These data, along with preliminary structural models of the scFvB80, both free and complexed with PSA, have allowed a working hypothesis to be developed regarding the interaction of PSA with its specific antibodies.

Acknowledgments

First and foremost, I would like to express my sincere gratitude to my supervisor, Dr. David S. Wishart, for his encouragement, enthusiasm and help. Without his intellectual direction and emotional support, I would never have been able to accomplish this work. His amiable personality, his broad range of knowledge and acute insight into problems have been indispensable assistance to my research in the past three years.

I would also wish to express my whole-hearted thanks to all my colleagues and fellow students I have had the pleasure of associating with over the past three years. In particular, I would like to thank Kathy Wang, Diane G. Jette and Yi Zhang for introducing me to the field of molecular biology and their assistance in the scFvB80 project (chapter 6). Kathy cloned the scFvB80 gene into *P. pastoris* and Diane measured the binding activity of scFvB80 using competitive ELISA (Figure 6.6). I would like to thank Alex M. Nip for the computer upkeep and in doing the computer simulations of the dependence of $^3J_{HNH\alpha}$ coupling constants on half-height linewidths (Figures 2.4 and 3.3). I would like to thank Gary Van Domselaar and Jane Nagel for their help in PSA epitope and PSA mimetope peptide synthesis (chapter 6). I would like to thank Wei Xia for her assistance on the scFvB80 expression and isotopically labeling expriments (chapter 6). Finally, I would like to thank Alan Gibbs for his assistance with the synthesis of the 22-mer CbnB2 N-terminal peptide (chapter 6).

I am very grateful to Dr. Brian D. Sykes and his group for their spectroscopic help over the past three years. I am also grateful to Dr. John C. Vederas and his group, Shengrong Chai, Nancy L. Fregeau Gallagher, Matthias E. Henz, Liang Z. Yan, for their cooperation and assistance in the CbnB2 project (chapter 5). In this project, Matthias synthesized LeuA and CbnB2 hexapeptide mimics and carried out kinetic studies of the disulfide bond redox reaction (Table 5.1); Shengrong and Nancy collected the initial NMR spectra of CbnB2 and some partial assignment; Yan designed and completed the carboamidomethylation experiments used to investigate the redox status of CbnB2 (Figure 5.1).

I particularly wish to express my deep gratitude to my wife and son, Yuhong and Pangpang, for their love, sacrifice, and encouragement over the past three years.

Finally, I would like to acknowledge the Pharmaceutical Manufacturers Association of Canada and Medical Research Council for their support of my PMAC-MRC postgraduate scholarship.

Table of Contents

Page

Chapter 1: Introduction 1

1.1 Protein structure determination by NMR 3

1.2 Measurement of $^3J_{HNHa}$ coupling constants 11

1.3 Bacteriophage T4 glutaredoxin 12

1.4 Carnobacteriocin B2 13

1.5 Anti-PSA single chain antibody fragment (scFvB80) 14

References 16

Chapter 2: A simple method to measure $^3J_{HNH\alpha}$ coupling constants from TOCSY and NOESY spectra 28

2.1 Introduction 28

2.2 Material and methods 30

2.3 Calculation of $^3J_{HNH\alpha}$ from $\Delta v_{1/2}$ 32

2.4 Results 36

2.5 Discussion 37

2.6 Conclusion 45

References 46

Chapter 3: Facile measurement of $^3J_{HNH\alpha}$ coupling constants from HMQC-J spectra 55

3.1 Introduction 55

3.2 Material and methods 56

3.3 Calculation of $^3J_{HNH\alpha}$ from $\Delta v_{1/2}$ 57

3.4 Results 59

3.5 Discussion 61

References 65

Chapter 4: NMR solution structures of oxidized and reduced bacteriophage T4 glutaredoxin 74
4.1 Introduction 74
4.2 Material and methods 76
4.3 Results and discussion 80
4.4 Conclusion 91
References 92

Chapter 5: The solution structure of carnobacteriocin B2: implication for structure-activity relationships among type IIa bacteriocins 117
5.1 Introduction 117
5.2 Material and methods 119
5.3 Results and discussion 125
5.4 Conclusion 135
References 136

Chapter 6: High level expression and characterization of an anti-PSA single chain antibody fragment in *P. pastoris* 155
6.1 Introduction 155
6.2 Material and methods 156
6.3 Results and discussion 164
6.4 Conclusion 171
References 173

Chapter 7: General discussion and conclusion 190
References 196

List of Tables

Page

Table 2.1 48

Table 2.2 49

Table 2.3 50

Table 3.1 68

Table 3.2 69

Table 3.3 70

Table 4.1 98

Table 4.2 100

Table 4.3 102

Table 4.4 103

Table 4.5 104

Table 4.6 105

Table 4.7 106

Table 5.1 141

Table 5.2 142

Table 5.3 144

Table 5.4 145

List of Figures

Page

Figure 1.1 21
Figure 1.2 21
Figure 1.3 22
Figure 1.4 23
Figure 1.5 23
Figure 1.6 24
Figure 1.7 24
Figure 1.8 25
Figure 1.9 25
Figure 1.10 26
Figure 1.11 27
Figure 2.1 51
Figure 2.2 52
Figure 2.3 53
Figure 2.4 54
Figure 3.1 71
Figure 3.2 72
Figure 3.3 73
Figure 3.4 74
Figure 4.1 107
Figure 4.2 108

Figure 4.3 109
Figure 4.4 110
Figure 4.5 111
Figure 4.6 112
Figure 4.7 113
Figure 4.8 114
Figure 4.9 115
Figure 4.10 116
Figure 5.1 146
Figure 5.2 147
Figure 4.7 113
Figure 4.8 114
Figure 4.9 115
Figure 4.10 116
Figure 5.1 146
Figure 5.2 147
Figure 5.3 148
Figure 5.4 149
Figure 5.5 150
Figure 5.6 151
Figure 5.7 152
Figure 5.8 153
Figure 5.9 154
Figure 6.1 177
Figure 6.2 178

Figure 6.3 179
Figure 6.4 180
Figure 6.5 181
Figure 6.6 182
Figure 6.7 183
Figure 6.8 184
Figure 6.9 185
Figure 6.10 186
Figure 6.11 187
Figure 6.12 188
Figure 6.13 189

List of Abbreviations

Amino acids:

A (Ala) = L-alanine
C (Cys) = L-cysteine
D (Asp) = L-aspartic acid
E (Glu) = L-glutanic acid
F (Phe) = L-phenylalanine
G (Gly) = L-glycine
H (His) = L-histidine
I (Ile) = L-isoleucine
K (Lys) = L-lysine
L (Leu) = L-leucine
M (Met) = L-methionine
N (Asn) = L-aspartic acid
P (Pro) = L-proline
Q (Gln) = L-glutamine
R (Arg) = L-arginine
S (Ser) = L-serine
T (Thr) = L-theonine
V (Val) = L-valine
W (Trp) = L-tryptophan
Y (Tyr) = L-tyrosine

Å = angstrom
BMGY = buffered minimal glycerol complex medium
BMMY = buffered minimal methanol complex medium
BPH = benign prostate hyperplasia
BSA = bovine serum albumin
CDR = complementarity-determining region
CD = circular dichroism
CSI = chemical shift index
CbnB2 = carnobacteriocin B2
DCM = dichloromethane

DMF = dimethylformamide

DMSO = dimethylsulfoxide

DQF-COSY = double quantum filtered correlation spectroscopy

DSS = 2,2-dimethyl-2-silapentane-5-sulfonic acid

DTT = dithiothreitol

E. coli = Esherichia coli

EDTA = ethylenediaminetetraacetic acid

ELISA = enzyme-linked immunosorbent assay

ES-MS = electrospray mass spectrometry

FMOC = 9-fluorenylmethyloxycarbonyl

Fab = antibody fragment

FID = free induction decay

Fv = antibody variable domain

Grx = glutaredoxin

HAMA = human anti-mouse antigen

HBTU = O-benzotriazole-N,N,N',N'-tetramethyluronium hexafluorophosphate

HMQC = heteronuclear multiple quantum correlation spectroscopy

HPLC = high-performance liquid chromatography

HRPO = horseradish peroxidase

HSQC = heteronuclear single-quantum coherence

IgG = immunoglobulin G

LAB = lactic acid bacteria

Leu A = leucocin A

MD = minimal dextrose medium

MM = minimal methanol medium

NMM = N-methylmorpholine

NMR = nuclear magnetic resonance

Mab = monoclonal antibody

MeCN = acetonitrile

NOE = nuclear Overhauser effect

NOESY = nuclear Overhauser effect spectroscopy

OD_{600} = optical density at 600 nm

PAGE = polyacrylamide gel electrophoresis

PBS = phosphate-buffered saline
PSA = prostate-specific antigen
RMSD = root mean square deviation
TFE = trifluoroethanol
TFA = trifluoroacetic acid
TMB = tetramethylbenzidine
TMS = tetramethylsilane
TOCSY = total correlation spectroscopy
TSP = 3-(trimethylsilyl)-propionate
Trx = thioredoxin
UV = ultraviolet
V_H = variable heavy chain
V_L = variable light chain
YPD = 1% yeast extract / 2% peptone / 2% dextrose
kD = kilodalton
scFv = single chain antibody variable domain

Chapter 1

Introduction

Over the past two decades, the rapid advances in structural biology have transformed the field of medicinal chemistry (i.e. drug design and drug discovery). Rather than treating protein drug targets as traditional "black boxes" we are now able to treat them as well-defined molecular entities that have a shape, a form and a specific site that can be activated or deactivated by a small molecule ligand. Improvements in structure determination methods and molecular rendering techniques have led to the development of "rational drug design"—a method of drug development that uses the structure of the protein target as an essential template in building or designing protein-specific small molecule ligands. Successful examples of rational drug design include the development of AChE inhibitors (to treat high blood pressure) (Gooch and Stennett, 1996; Cho, 1996), HIV protease inhibitors (to treat AIDS) (DesJarlais et al., 1990; Olson and Goodsell, 1998; Kiso, 1998) and antimitotic agents such as, epothilones and clerocidins (Amat et al., 1998).

One area of structural biology that is playing an increasingly important role in rational drug design is Nuclear Magnetic Resonance (NMR) spectroscopy. Indeed, NMR has emerged as the most powerful technique for determining the three-dimensional solution structures of pharmaceutically important peptides and proteins (Fairlie et al., 1998; Gmeiner, 1998). Many technological improvements in NMR spectrometer technology (increased field strength, faster computers, improved probe sensitivity, gradient technology, etc.) coupled with improved pulse sequences for performing highly specific heteronuclear multidimensional experiments (Kay and Gardner, 1997; Kay, 1995) have allowed proteins and protein complexes as large as 67 kilodalton to be fully assigned and characterized by NMR (Shan et al., 1996). Despite these improvements in NMR technology, there are a number of areas of biomolecular NMR spectroscopy still in need of improvement. In particular, the requirement for abundant quantities of isotopically labeled protein continues to limit the types and numbers of pharmaceutically interesting proteins that can be studied by NMR. Indeed the development of better protein expression systems and improved isotopic labeling techniques still lags far behind the rapid developments in NMR spectrometer technology (Jansson et al., 1996). Another area of biomolecular NMR spectroscopy that seems to be neglected concerns the use of chemical shift and coupling constant data in the determination of protein structures. Even today, most NMR spectroscopists are still bound to the "tried and true" methods of using only NOE data to obtain restraints for protein structure determination. This largely arises from the difficulty in interpreting or measuring other scalar quantities (such as chemical shifts or $^3J_{HNH\alpha}$ coupling constants) that, in principle, could provide as much or even more structural detail as NOE measurements. It is these and other perceived limitations in NMR methodology that has served as the motivation for the work described in this thesis.

Specifically, this dissertation is concerned with the development of novel methods to improve the potential of NMR spectroscopy in characterizing pharmaceutically useful proteins. In particular, it describes several methods that address the previously mentioned problems concerning protein expression and labeling for NMR experiments. This thesis also describes two new and very simple methods for determining $^3J_{HNH\alpha}$ coupling constants for medium-to-large (> 10 kDa) protein molecules. Using these new techniques it will be shown how they can be used in the production and/or characterization of three proteins of pharmaceutical interest: (1) carnobacteriocin B2 (48 residues), (2) bacteriophage T4 glutaredoxin (87 residues) and (3) an anti-PSA single chain antibody fragments (called scFvB80, 261 residues). Chapters 2 and 3 describe two new methods for quantitatively measuring $^3J_{HNH\alpha}$ coupling constants in peptides and proteins. The difficulties experienced in implementing previously published techniques to measure the $^3J_{HNH\alpha}$ coupling constants for T4 glutaredoxin were the direct motivation for the generation of these two simple, yet powerful methods. Chapters 4 and 5 describe the structure determination, by NMR techniques, of T4 glutaredoxin and carnobacteriocin B2. Chapter 6 describes the production and isotopic labeling of a difficult-to-express protein (scFvB80) using a simple eukaryotic expression vehicle (*P. pastoris*) and a novel isotopic labeling protocol. In the same chapter, the study of the interaction between scFvB80 and PSA as well as PSA peptide epitopes/mimetopes by homology modeling, docking and ^1H one-dimensional NMR techniques are also reported.

Because this thesis is prepared in a paper format, each chapter has its own detailed introduction. Rather than attempt to repeat the contents of each chapter, the following discussion will be limited to some basic background information not covered in subsequent chapters.

1.1. Protein structure determination by NMR

Physical principles of NMR. Nuclear magnetic resonance spectroscopy deals with the interaction between the magnetic moments of atomic nuclei and a magnetic field (Kessler, et el., 1988). The magnetic moment of the nucleus is associated with the nuclear spin *I*. From quantum mechanics, each nuclear particle (e.g., proton or neutron) has a spin value of I=1/2. The combination of multiple particles in the nucleus results in an overall spin value that is specific for each atomic isotope. Those isotopes with an even number of protons and neutrons will have zero magnetic spin (e.g., ^{4}He, ^{12}C and ^{16}O). An odd number of protons and an even number of neutrons or an odd number of neutrons and an even number of protons (e.g., ^{1}H, ^{13}C, ^{15}N, or ^{19}F) result in an overall 1/2 or a multiple of 1/2 spin. Those isotopes with odd numbers of both protons and neutrons (e.g., ^{2}H or ^{14}N) have more complex spin states and are less suitable for direct NMR observation in macromolecules. Each of the four most abundant elements in biological materials (H, C, N, and O) have at least one natural occurring isotope with non-zero nuclear spin. Therefore, each of these elements is, in principle, observable in an NMR experiment. The naturally occurring isotope of hydrogen, ^{1}H, is present at >99% abundance and forms the basis of most of the experiments described in this thesis. Other important NMR-active isotopes include ^{13}C and ^{15}N present at 1.1% and 0.4% natural abundance, respectively. The low natural abundance of these two isotopes makes their observation difficult for most natural or biological products. These two nuclei, however, are very extensively used to characterize larger (>10 kDa) proteins.

In the presence of an external magnetic field, the spin angular momentum of atoms with overall non-zero spin will undergo a cone-shaped rotational motion called precession (Figure 1.1). The rate of precession for each isotope is referred to as the Larmor frequency (ω_0), which is dependent of the strength of the external field (B_0) and the intrinsic properties of the nucleus reflected in its gyromagnetic ratio γ. The

Larmor frequency may be represented by:

$$(1.1) \qquad \omega_0 = - \gamma B_0$$

Each magnetic nucleus has 2*I*+1 possible orientations and corresponding energy levels with respect to the external magnetic field. Thus, for nuclei with *I*=1/2, there are two possible orientations in the magnetic field: parallel and antiparallel, which corresponds to two energy levels (Figure 1.2). The energy difference between these two orientations is directly proportional to the strength of the magnetic field.

$$(1.2) \qquad \Delta E = \gamma h B_0 / 2\pi$$

Where h is Planck's constant.

At any given magnetic field, a nucleus can jump from one energy level to the other by absorbing or emitting a discrete amount of energy at its Larmor frequency. At room temperature, the two energy states are unequally populated and the population ratio can be expressed by the Boltzmann equation:

$$(1.3) \qquad N_\beta / N_\alpha = \exp(-\Delta E / kT)$$

Where N_α and N_β are the population of the lower and upper states respectively.

This population difference generates a net magnetization (M), which aligns with the external magnetic field and will remain static unless the system is disturbed in some way. If a magnetic pulse is applied for a short period of time in such a way that it produces a second magnetic field (B_1) perpendicular to the static field B_0, it will drive M away from it equilibrium position by a so-called "flip angle", depending on the time period and the field strength of B_1. A 90° pulse is therefore the time it takes a particular field strength to rotate the equilibrium magnetization 90° with respect the its equilibrium direction (the direction of the applied magnetic field, usually set along the Z-axis). Having a metal coil in the XY-plane allows the recording of the oscillating current generated by the precessing magnetization. The precessing magnetization eventually goes back to equilibrium, with the XY magnetization slowly fading and the Z magnetization growing. This oscillating magnetic field can be detected by a coil, converted to an electrical signal and recorded. This signal is called the free-induction decay (FID). Using a mathematical operation called Fourier transform; the information can then be converted from the FID (time domain) into the more familiar spectrum (the frequency domain). The whole process is illustrated in Figure 1.3.

Chemical shifts. The magnetic field at the nucleus is not exactly equal to the applied magnetic field: electrons around the nucleus shield it from the applied magnetic field. The greater the electron density about a nucleus, the more that nucleus is "shielded". Because of this, a stronger magnetic field or higher frequency is required to bring the nucleus into resonance. The degree of shielding depends on whether a neighboring chemical group withdraws electron density from the nucleus. This electron withdrawal may occur either by induction or by a resonance effect. The difference between the applied magnetic field and the field at the nucleus is called nuclear shielding. The chemical shift is defined as the nuclear shielding divided by the applied magnetic field. Chemical shifts are measured relative to reference compounds.

TMS (tetramethylsilan), TSP (3-(trimethylsilyl)-propionate) and DSS (2,2'-dimethyl-2-silapentane-5-sulfonate) are commonly used (zero-point) standards for ^1H and ^{13}C spectra. Because of the problems of solubility with TMS and pH sensitivity with TSP, DSS represents the best current choice as internal ^1H and ^{13}C zero-point standard (Wishart et al., 1995). Liquid ammonia is commonly used as an ^{15}N external zero-point standard in biomolecular NMR while TFA (trifluoroacetic acid) is the commonly used standard for ^{19}F chemical shifts (Wishart and Nip, 1998). The hemical shift is defined as the difference between the field strength at which the particular nucleus absorbs with respect to the nuclei of a given chemical shift standard. The scale is most commonly used as a measure for reporting chemical shifts.

$$(1.4) \qquad \delta = \frac{\textbf{Shift from standard (Hz) x } 10^6}{\textbf{Spectrometer frequency (Hz)}} \ \textbf{(ppm)}$$

Coupling constants. Nuclei experiencing the same chemical environment or chemical shift are called equivalent. Those nuclei experiencing different environment or having different chemical shifts are nonequivalent. Nuclei, when close to one another, exert an effect on each other's effective magnetic field. This effect manifests itself on the NMR spectrum when the nuclei are nonequivalent. If the separation between non-equivalent nuclei is less than or equal to three (some times four) bond lengths, this effect is observable. It is called spin-spin coupling or J coupling. J couplings between pairs of protons separated by three or fewer covalent bonds can be measured. The value of the three-bond vicinal coupling constant (^3J) contains information about the intervening torsion angle. This relationship has the form (Karplus, 1959):

$$(1.5) \qquad {}^3J = A \cos(\theta) + B \cos^2(\theta) + C$$

Where A, B, and C are empirically derived constants for each type of coupling constant.

Coupling constants can provide very important structural information about dihedral angles in proteins (Bystrov, 1976; Pardi et al., 1984, Wang and Bax, 1995 and 1996). The three-bond coupling constant between the intra-residual alpha and amide protons ($^3J_{HNH\alpha}$) is most useful for secondary and tertiary structure determinations as it can be directly related to the backbone dihedral angle phi (Figures 1.4 and 1.5).

Protein structure determination by NMR. Since the early 1980s, high resolution NMR spectroscopy has become the most important tool for qualitatively and/or quantitatively determining protein solution structures (Wuthrich, 1990). Basically this process can be divided into three steps:

Step 1. Chemical shift and NOE assignment. The key to solving a protein structure by NMR is to assign each resonance to a specific amino acid and then to identify through-space NOE interactions between the assigned resonances. Basically, only two types of NMR experiments are used to sequentially assign a protein: those identifying through-bond connectivity, e.g. COSY (Aue et al., 1976; Rance et al., 1983), TOCSY (Griesinger et al., 1988), HNCOCA (Grzesiek and Bax, 1992) etc., and those identifying through-space correlation, e.g., NOESY (Bodenhausen et al., 1984). The overall strategy of the assignment can be separated into two steps: (1) identification of spin systems using through-bond connectivities, and (2) sequential assignments of these spin systems using through-space connectivities (i.e. the observed NOEs between adjacent residues). This second step usually begins with unique residues and unique pairs of residues for which the NOEs give the highest probability of being sequential.

For the three proteins described in this thesis, homonuclear ^{1}H two-dimensional (2D) experiments (Jeener et al., 1979; Aue et al., 1976) were primarily used for structural analysis and determination. 2D NMR spreads the spectral information into two frequency dimensions, thus improving the effective resolution and sensitivity of the NMR experiment. All 2D NMR pulse sequences may be described in terms of four successive time periods: 1) preparation, 2) evolution, and 3) mixing, and 4) detection. The preparation period usually consists of a delay time (during which thermal equilibrium or a steady state is attained) followed by one or more radio frequency pulses that create the desired coherence. During the evolution period, the coherence of the system is frequency-labeled in a manner that depends on the length of the evolution period. The mixing period may include one or more radio frequency (RF) pulses and delay intervals. During the mixing period (τ_m), information is relayed between spins by through-bond magnetization transfer, so that connections between nuclei labeled in the evolution period and nuclei detected during the detection period are established. During the detection period the system evolves further and the resulting FID is recorded. The pulse sequence for a conventional Total Correlation Spectroscopy experiment (TOCSY; Braunschweiler and Ernst, 1983), which permits correlation of all protons within a scalar coupling network, is shown in Figure 1.6. An example of a homonuclear TOCSY spectrum is shown in Figure 1.7.

The most useful NMR experiment for the structural NMR spectroscopist involves, without any doubt, the through-space correlation experiment. This correlation, named the Nuclear Overhauser effect (NOE) (Noggle et al., 1971), occurs between nuclei that are close in space. In simplest terms, the magnitude of the NOE is proportional to $1/r^6$, where r is the distance between two nuclei. The sign and intensity of the NOE also depends on the correlation time describing the motion of the inerproton bond vector. Due to this relation, the NOE fades quickly with distance, and NOE's are not normally observed between protons that are more than 5 Å apart. Nevertheless, the NOE is very sensitive to internuclear distances and therefore may be used to estimate proton-proton distances. Observed NOE's may have either a positive or negative phase. If the molecular tumbling rate is fast (small molecules) NOE's are negative; if the tumbling rate is slow (large molecules) the signals are positive. The pulse sequence for a conventional 2D Nuclear Overhauser Effect Spectroscopy experiment (NOESY; Jeener et al., 1979) is shown in Figure 1.8. In the NOESY experiment, evolution period consists of two 90° pulses followed by a mixing time (τ_m) during which cross-relaxation involving longitudinal magnetization occurs. An example of a NOESY spectrum is shown in Figure 1.9.

Step 2. Secondary structure determination. NMR spectroscopy is perhaps the most powerful and certainly the most accurate spectroscopic method of secondary structure determination. Currently four major NMR parameters, e.g., short and medium-range NOEs (Wüthrich, 1986), $^3J_{HNH\alpha}$ coupling constants (Wuthrich, 1986), slowly exchanging amide protons (Hruby, 1974), and chemical shifts (Wishart et al., 1991a, 1991b, 1992 and 1995; Wishart and Sykes, 1994; Osapay and Case, 1994; Williamson and Asakura, 1992) can be used to evaluate the existence of secondary structure. Combinations of these four parameters will provide a reliable indication of the location of protein secondary structural segments.

A number of short (< 5 Å) inter-proton distances are fairly unique to certain secondary structural elements. For example, alpha helices are characterized by short distances between certain protons on sequentially neighboring residues (e.g., between backbone amide protons (d_{NN}) as well as between beta protons of residue i and the amide protons of residue i+1 ($d_{\beta N}$). Helical conformations result in short distances between the alpha proton of residue i and the amide proton of residues i+2; i+3 and i+4. These i+2, i+3, and i+4 NOEs are collectively referred to as medium-range NOEs while NOEs connecting residues separated by more than 5 residues are referred to as long-range NOEs. Extended conformations e.g., β-strands, are characterized by short sequential distances between the α-proton and the adjacent amide proton ($d_{\alpha N}$). The formation of β–sheets also result in short distances between protons on adjacent -strands (e.g., $d_{\alpha\alpha}$ and $d_{\alpha N}$) (Wuthrich, 1986).

$^3J_{HNH\alpha}$ coupling constants are also very useful tools for secondary structure determinations. For example, helical and extended conformations have very different values for phi (-60° and -120°, respectively) which result in measurable differences in $^3J_{HNH\alpha}$ (Wuthrich, 1986) right-handed alpha helix, phi = -57°, $^3J_{HNH\alpha} = 3.9$ Hz

right handed 3-10 helix, phi = -60°, $^3J_{HNH\alpha} = 4.2$ Hz

antiparallel beta sheet, phi = -139°, $^3J_{HNH\alpha} = 8.9$ Hz

parallel beta sheet, phi = -119°, $^3J_{HNH\alpha} = 9.7$ Hz

left-handed alpha helix, phi = 57°, $^3J_{HNH\alpha} = 6.9$ Hz

Regular hydrogen-bonded secondary structures "protect" amide protons from solvent H_2O exchange. Although nearly 75% of a polypeptide's amide protons are typically involved in hydrogen bonds (Baker and Hubbard, 1984), those in regular secondary structures appear to be longer-lived. Therefore amide proton exchange rates provide qualitative important information about the location and stability of secondary structures. For example, continuous stretches of four or more slowly exchanging amide protons can often indicate the existence of a helix. Alternating stretches of slowly and rapidly exchange amide protons can indicate the existence of an exterior β–strand.

Because the chemical shift of a nucleus is sensitive to its environment, it should also contain structural information. Correlations between chemical shift tendencies and secondary structures have been identified (Wishart et al., 1992). The alpha protons as well as the Cα, Cβ and CO chemical shifts of all 20 natural amino acids have been shown to have a strong correlation with secondary structure. Wishart et al. (1991 and 1992) described a simple method for secondary structure determination by analyzing the difference between the alpha proton chemical shift for each residue and that reported for the same residue in a "random coil" conformation. Helical segments have groupings of alpha protons whose chemical shifts are consistently less than the random coil values whereas beta strands have values consistently greater. In this way the identification and location of α-helix and β-strand segments are possible and quite reliable.

Step 3. Tertiary structure calculation. Two computational strategies: 1) distance geometry (Havel and Wuthrich, 1985; Williamson et al., 1985); and 2) restrained dynamically simulated annealing (Nilges et al., 1988) are commonly used to generate three dimensional protein structures from NMR data. In distance geometry, the molecule is represented as a series of inter-atomic distances. The number of inter-atomic distances for a molecule consisting N atoms is equal to N(N-1)/2. This will yield an N x N matrix of distances, called a distance matrix. The task of distance geometry is to convert the distance matrix into the traditional Cartesian coordinates describing the atomic structure of the molecule of interest. To do so, two distance matrices, a lower bound matrix (containing the minimum inter-atomic distances) and an upper bound matrix (containing maximum distances), are required. These two matrices are filled with upper and lower-bound distance data derived from NOEs. Diagonalization of these matrices followed by further manipulation can yield Cartesian coordinates of the protein.

In contrast to distance geometry, the molecular dynamics approach assigns velocities to all of the atoms, then carries out classical trajectory calculations using an energy function which is supplemented by restraint terms derived from experimental data (i.e. NOE derived proton-proton distance and dihedral torsion angle restraints). During such a run, various conformations can be extracted and energy minimized. Simulated annealing is a molecular dynamics simulation technique where the temperature of the system is artificially increased to approximately 10000K. At this elevated temperature the molecule is able to traverse all potential barriers. During the annealing phase, the temperature of the system is then gradually reduced during the course of the simulation. The heating/cooling process in repeated until no new energy minima can be found and the resulting structures satisfies the experimental data with an acceptable error. Because simulated annealing approach usually carries out trajectory calculations using just the bonding, hard sphere van der Waals and distance constraint terms in the "potential" function (eliminating attractive dispersive and electrostatic terms, which are too slow to calculate), this process is much faster than full molecular dynamics calculations. Furthermore, the initial structures used to begin this process seem not to be particularly crucial as extended structures, crudely embedded structures, and even randomly assigned coordinates can give comparably good results.

Regardless of the strategy applied, successful structures generated from these calculations should satisfy all the experimentally measured restraints, covalent geometry, and non-bonded contacts. For NMR structure determination the experimental restraints include NOE derived distance constraints and dihedral angle restraints. Recently coupling constant restraints (Garrett et al., 1994) and chemical shift restraints (Kuszewski et al., 1995) have also been adopted to directly refine NMR structures.

Typically, 20-60 structures must be calculated in order to fully sample the allowed conformational space so that a given structure agrees with the experimental data. A detailed procedure using the simulated annealing protocol implemented in X-PLOR (Brunger, 1992) to generate an NMR structures of T4 glutaredoxin and carnobacteriocin B2 are described in Chapters 3 and 4.

1.2. Measurement of $^3J_{HNH\alpha}$ coupling constants

Coupling constant measurements are playing an increasingly important role in the determination and refinement of peptide and protein structures (Bax et al., 1994). Their rise in importance has had much to do with the recent introduction of innovative experimental and analytical methods that have facilitated the extraction of these otherwise difficult-to-measure parameters. A variety of ingenious curve-fitting techniques is now available for determining $^3J_{HNH\alpha}$ coupling constants from homonuclear DQF-COSY spectra (Pardi et al., 1984; Kim and Prestegard, 1989; Smith et al., 1991), from NOESY spectra (Szyperski et al., 1992) or a combination of DQF-COSY and NOESY spectra (Ludvigsen et al., 1991). Each of these methods is based on computationally intensive procedures that attempt to fit theoretical curves to anti-phase or in-phase doublets.

With the widespread use of isotopic labels in biological NMR, a number of elegant experimental methods has also been developed to measure $^3J_{HNH\alpha}$ values from ^1H-^{15}N heteronuclear experiments. These include the HMQC-J experiment (Kay and Bax, 1990), the J-modulated [^{15}N,^1H] COSY experiment (Billeter et al., 1992), the HNCA E. COSY (Weisemann et al., 1994) and the HNHA experiment (Vuister and Bax, 1993). By combining spectral data from these experiments with experiment-specific computer curve-fitting routines (Kay and Bax, 1990; Goodgame and Greer, 1993) or peak integration routines (Vuister and Bax, 1993; Billeter et al., 1992), it is possible to extract relatively accurate $^3J_{HNH\alpha}$ coupling constants. Furthermore, the use of isotopically labeled proteins, with their inherently greater chemical shift dispersion, now permits the measurement of $^3J_{HNH\alpha}$ values of much larger proteins (up to 18 kDa - Billeter et al., 1992).

Despite the arsenal of experimental and computational techniques available to quantitatively determine $^3J_{HNH\alpha}$ coupling constants, it is still quite rare to see $^3J_{HNH\alpha}$ coupling constants reported in the literature. In fact, coupling constant measurements often require special hardware or special computer programs, or the implementation of a complex pulse sequence.

Chapters 2 and 3 in this thesis describe the development of two new methods to quantitatively measure $^3J_{HNH\alpha}$ coupling constants, based on the simple observation that $^3J_{HNH\alpha}$ coupling constants are linearly related to the linewidth at half-height ($\Delta\nu_{1/2}$) of in-phase TOCSY, NOESY and HMQC-J amide crosspeaks. Tests on nearly twenty different proteins and peptides show that these new methods offer several significant advantages over most other published techniques. These two methods were also successfully applied to the structure determination of bacteriophage T4 glutaredoxin and carnobacteriocin B2.

1.3. Bacteriophage T4 Glutaredoxin

Thioredoxins and glutaredoxins are small ubiquitous proteins containing an active site with a redox-active disulfide (Holmgren, 1985). They function in electron transfer reactions via the reversible oxidation of two vicinal protein-SH groups to form a disulfide bridge. Both types of protein are involved in ribonucleotide reduction and other NADPH dependent reactions (Holmgren, 1989). Glutaredoxins also have other important biological functions, e.g., catalyzing the glutathione dependent reduction of dehydroascorbate to ascorbate (Wells et al., 1990; Martensson and Meister, 1991), reactivating DNA-binding activity (Bandyopadhyay et al., 1998), and regulating and/or maintaining HIV-1 protease activity *in vitro* (Davis et al., 1997).

Thioredoxin and glutaredoxin form a functionally and structurally related super-family (Holmgren and Bjornstedt, 1995). Bacteriophage T4 glutaredoxin (T4 Grx) seems to perform the hybrid functions of both thioredoxin and glutaredoxin. Interestingly, the lack of sequence similarity to any other glutaredoxins or thioredoxins indicates that this protein may fall in between these two families (Figure 1.10). Therefore, structural studies of this "missing-link" protein could offer some unique insights into the structure/function of both glutaredoxins and thioredoxins. Several X-ray crystal structures of T4 glutaredoxin and its mutant protein have been reported, e.g., for the oxidized form at 2.8 Å resolution (Soderberg et al., 1978) and at 2.0 Å resolution (Eklund et al., 1992), for a mutant protein (Val15 to Gly and Tyr16 to Pro) in the oxidized form at 1.45 Å resolution (Eklund et al., 1992) and a mutant in the reduced form at 2.4 Å resolution (Ingelman et al., 1995). More detailed information about this protein is given in Chapter 4, which describes the determination of the solution structures of wildtype T4 glutaredoxin in both its oxidized and reduced forms.

1.4. Carnobacteriocin B2

Chapter 5 reports on the NMR structure of Carnobacteriocin B2 (CbnB2) in a 90% TFE/water mixture. CbnB2 is a type IIa bacteriocin produced by *Carnobacterium piscicola* LV17B, a Lactobacillus-type organism originally isolated from refrigerated meats (Quadri et al., 1994 and 1995). CbnB2 targets the cytoplasmic membrane of sensitive cells and causes dissipation of the proton-motive force with concomitant leakage of intracellular components (Quadri et al., 1997). The fact that CbnB2 (along with other members of this family of peptides) is non-toxic to humans yet bactericidal to many microbes that cause food spoilage and disease has led to further investigation of CbnB2 as nontoxic food-preserving agent (Vandenbergh, 1993). Although a large number of studies on antimicrobial peptides have been carried out (Tichaczek et al, 1994; Holck et al, 1994; Bukhtiyarova et al., 1994; Nissen-Meyer and Nes, 1997; Casaus et al., 1997), only one structure of a type IIa bacteriocin, Leucocin A (LeuA) has been reported so far (Fregeau et al., 1997). Cbn B2 has considerable sequence similarity to LeuA as well as to other bacteriocins produced by lactic acid bacteria, especially near the N-terminus, where a characteristic motif Tyr-Gly-Asn-Gly-Val (YGNGV) is highly conserved (Quadri et al., 1994 and 1997). As described in Chapter 5, CbnB2 and LeuA, despite sharing considerable sequence similarity, have very different structures in their N-terminal regions. This may have interesting implications regarding the structure-activity relationships being developed for these peptides/proteins.

1.5. Anti-PSA single chain antibody fragment (scFvB80)

An antibody (Ab) is a protein employed by the immune system to seek out and help destroy foreign material in the body. The basic unit of a natural antibody (or immunoglobulin, Ig) consists of two identical heavy-chain and two identical light-chain polypeptides (Kuby, 1994). At the protein level each of these chains folded into discrete domains (Figure 1.11). The N-termini of both chains exhibit considerable sequence variability and, hence are called "variable" or V regions. The V-regions of one heavy and one light chain associate to form the antigen-binding site of an antibody. The C-terminal ends of both heavy and light chains are more conserved in sequence and constitute the "constant" region. The whole antibody is a Y-shaped protein, ~150 kDa in size. Since the discovery of hybridoma technology in the mid-1970's, monoclonal antibodies (Mabs) have gained increasing importance as reagents in diagnostic and therapeutic medicine (Kohler and Milstein, 1975; Winter and Milstein, 1991). Human monoclonal antibodies would be preferable for most medical applications, but their production has been hampered by considerable technical difficulties. Recombinant murine antibodies provide an alternative (Verhoeyen and Riechmann, 1988). However, the fact that intact murine monoclonal antibodies cause unwanted human immune responses has limited their use in therapeutic or *in vivo* diagnostic applications (Shawler et al., 1985). Given that most of the host immune response comes from the constant region of a foreign antibody, the removal of this portion could simultaneously reduce the HAMA (human anti-mouse antigen) reaction while simplifying the problem of antibody production. It has been proposed that the development of single-chain antibody-based agents (scFv, i.e. antibodies without the constant region) could be an effective way to reduce unwanted immune responses. A single chain antibody is a genetically engineered recombinant fusion protein composed of only a heavy chain (V_H) and a light chain (V_L) variable domain connected by an artificial peptide linker (Bird et al., 1988). In addition to its reduced HAMA activity, this minimalist antibody can offer increased accessibility to tumor cells *in vivo*, and may be of considerable utility in targeted drug, radionuclide or hormone delivery systems (Yokota et al., 1992; Colcher et al., 1990).

Chapter 6 reports the cloning, expression, purification, isotopic labeling, and characterization of an anti-PSA single chain antibody (designated as scFvB80). Prostate specific antigen (PSA), an important tumor-associated marker (Hostetler et al., 1996; Mettlin et al., 1993), is a 33-kDa single-chain glycoprotein produced by the prostate epithelium. It is normally found in low concentrations (<2.5 μg/L) in male blood plasma but its concentration can be substantially increased (5-100X) in conjunction with prostate cancer, benign prostate hyperplasia (BPH), and surgical trauma to the prostate (Christensson et al., 1990; Luderer et al., 1995). Monoclonal antibody (Mab)-based RIAs and ELISAs are now regularly used to measure PSA levels in serum. The results are used to screen for prostate cancer and to monitor patients during cancer treatment (Armbruster, 1993; Ploch and Brawer, 1994; Schellhammer et al., 1993). B80.3, a murine IgG anti-PSA monoclonal antibody (developed by Biomira Inc), has been shown to have a very high affinity (K_d=1X10^{-9} M) to PSA (Jette et al., 1996; Kreutz and Suresh, 1997). More importantly, its gene sequence is known; making it one of only two anti-PSA antibodies that has been fully cloned and sequenced. The central objective of this project was to use scFvB80 as a mimic of the B80 Mab to study the interaction between PSA and a PSA-specific antibody (Figure 1.12) with the eventual intent to use this information to aid in the development of improved diagnostic tests and/or prostate cancer therapies. As a first step towards studying these interactions, models of PSA bound to scFvB80 were generated using homology modeling and computer-aided docking. In addition, the interaction between scFvB80 and various PSA peptide epitopes and mimetopes was studied experimentally using one-dimensional ^{1}H NMR.

References

Amat, L., Robert, D., Besalu, E. and Carbo-Dorca , R. (1998) *Journal of Chemical Information and Computer Sciences* 38, 624-31.

Armbruster, D. A. (1993) *Clin. Chem.* 39, 181-195.

Aue, W. P., Bartholdi, E., Ernst, R. R. (1976) *J. Chem. Phys.* 64, 2229-2246.

Baker, E. N. and Hubbard, R. E. (1984) *Prog. Biophys. Molec. Biol.* 44, 97-179.

Bandyopadhyay, S., Starke, D. W., Mieyal, J. J. and Gronostajski, R. M. (1998) *J. Biol. Chem.* 273, 392-397.

Bax, A., Vuister, G. W., Grzesiek, S., Delaglio, F., Wang, A. C., Tschudin, R. and Zhu, G. (1994) *Meth. Enzymol.* 239, 79-105.

Billeter, M., Neri, D., Otting, G., Qian, Y. Q. and Wuthrich, K. (1992) *J. Biomol. NMR* 2, 257-274.

Bird, R. E., Hardman, K. D., Jacobson, J. W., Johnson, S., Kaufman, B. M., Lee, S-M., Lee, T., Pope, S. H., Riordan, G. S. and Whitlow, M. (1988) *Science* 242, 423-426.

Bodenhausen, G., Kogler, H. and Ernst, R. R. (1984) *J. Magn. Reson.* 58, 370-388.

Braunschweiler, L. and Ernst, R. R. (1983) *J. Magn. Reson.* 53, 521-528.

Brunger, A. T. (1992) X-PLOR. *A System for X-ray Crystallography and NMR*, Yale University Press, New haven, CT.

Bukhtiyarova, M., Yang, R. and Ray, B. (1994) *Appl. Environ. Microbiol.* 60, 3405-3408.

Bystrow, V. F. (1976) *Prog. NMR Spectrosc.* 10, 41-81.

Casaus, P., Nilsen, T., Cintas, L.M., Nes, I. F., Hernandez, P. E., Holo, H. (1997) *Microbiology* 143, 2287-2294.

Cho, S. J., Garsia, M. L., Bier, J. and Tropsha, A. (1996) *J. Med. Chem.* 39, 5064-71.

Christensson, A., Laurell, C. B. and Lilja, H. (1990) *Eur. J. Biochem.* 194, 755-763.

Colcher, D. R., Bird, R. E., Roselli, M., Hardman, K. D., Johnson, S., Pope, S., Dodd, S. W., Pantoliano, M. W., Milenic, D. E. and Schlom, J. (1990) *J. Natl. Cancer Inst.* 82, 1191-1197.

Davis, D. A., Newcomb, F. M., Starke, D. W., Ott, D. E., Mieyal, J. J. and

Yarchoan, R. (1997) *J. Biol. Chem.* 272, 25935-25940.

DesJarlais, R. L., Seibel, G. L., Kuntz, I. D., Furth, P. S., Alvarez, J. C., Ortiz, P. R., DeCamp, D. L., Babe, L. M. and Craik, C. S. (1990) *Proc. Natl. Acad. Sci. U.S.A.* 87, 6644-6648.

Eklund, H., Ingelman, M., Soderberg, B. O., Uhlin, T., Nordlund, P., Nikkola, M., Sonnersam, U., Joelson, T. and Petratos, K. (1992) *J. Mol. Biol.* 228, 596-618.

Fairlie, D. P., West, M. L. and Wong, A. K. (1998) *Current Medicinal Chemistry* 5, 29-62.

Fregeau Gallagher, N. L., Sailer M., Niemczura, W. P., Nakashima, T. T., Stiles, M. E. and Vederas, J. C. (1997) *Biochemistry* 36, 15062-15072.

Garrett, D. S., Kuszewski, J., Hancock, T. J., Lodi, P. J. Vuister, G. W., Gronenborn, A. M. and Clore, G. M. (1994) *J. Magn. Reson. B* 104, 99-103.

Gooch, M. D. and Stennett, D. J. (1996) *American Journal of Health-System Pharmacy* 53, 1545-57.

Goodgame, M. M. and Greer, S. M. (1993) *J. Magn. Reson. A* 102, 246-248.

Gmeiner, W. H. (1998) *Current Medicinal Chemistry* 5, 115-35.

Griesinger, C., Otting, G., Wuthrich, K. and Ernst, R. R. (1988) *J. Am. Chem. Soc.* 110, 7870-7872.

Grzesiek, S. and Bax, A. (1992) *J. Magn. Reson.* 96: 432-440.

Havel, T. F. and Wuthrich, K. (1985) *J. Mol. Biol.* 182, 281-294.

Holck, A., Axelsson, L., Birkeland, S. E., Aukrust, T. and Blom, H. (1992) *J. Gen. Microbiol.* 138, 2715-2720.

Holmgren, A. (1985) *Annu. Rev. Biochem.* 54, 237-271.

Holmgren, A. (1989) *J. Biol. Chem.* 264, 13963-13966.

Holmgren, A. and Bjornstedt, M. (1995) *Methods Enzymol.* 252, 199-208.

Hostetler, R. M., Mandel, I. G. and Marshburn, J. (1996) *Med .Clin. North. Am.* 80, 83-98.

Hruby, V. J. (1974) *Chemistry and Biochemistry of Amino Acids, Peptides and Proteins* Weinstein, B. Ed., Vol. 3, New York.

Ingelman, M., Nordlund, P. and Eklund, H. (1995) *FEBS Letters* 370, 209-211.

Jansson, M., Li, Y. C., Jendeberg, L., Anderson, S., Montelione, B. T. and Nilsson, B. (1996) *J. Biomol. NMR* 7 131-141.

Jeener, J., meier, B. H., Bachmann, P. and Ernst, R. R. (1979) *J. Chem. Phys.* 71, 4546-4553.

Jette, D. C., Kreutz, F. T., Malcolm, B. A., Wishart, D. S., Noujaim, A. A. and Suresh, M.R. (1996) *Clin. Chem.* 42, 1961-1969.

Karplus, M. (1959) *J. Chem. Phys.* 10, 11-15.

Kay, L. E. (1995) *Progress in Biophysics and Molecular Biology* 63, 277-99.

Kay, L. E. and Bax, A. (1990) *J. Magn. Reson.* 86, 110-126.

Kay, L. E. and Gardner, K. H. (1997) *Current Opinion in Structural Biology* 7, 722-31.

Kessler, H., Gehrke, M. and Griesinger, C. (1988) *Angew. Chem. Int. Ed. Engl.* 27, 490-536.

Kim, Y. and Prestegard, J. H. (1989) *J. Magn. Reson.* 84, 9-13.

Kiso Y. (1998) *Protein, Nucleic Acid, Enzyme*. 43, 725-733.

Kohler, G. and Milstein, C. (1975) *Nature* 256, 495-497.

Kreutz, F. T. and Suresh, M. R. (1997) *Clin. Chem.* 43, 649-656.

Kuby, J. (1994) *Immunology*, 2nd Edition, New York.

Kuszewski, J., Gronenborn, A. M. and Clore, G. M. (1995) *J. Magn. Reson. B* 107, 293-297.

Luderer, A. A., Chen, Y. T., Soriano, T. F., Kramp, W. J., Carlson, G., Cuny, C., Sharp, T., Smith, W., Petteway, J. and Brawer, M. K. (1995) *Urology* 46, 187-194.

Ludvigsen, S., Andersen, K. V. and Poulsen, F. M. (1991) *J. Mol. Biol.* 217, 731-736.

Martensson, J. and Meister, A. (1991) *Proc. Natl. Acad. Sci. U.S.A.* 88, 4656-4660.

Mettlin, C., Murphy, G. P., Ray, P., Shanberg, A., Toi, A., Chesley, A., Babaian R., Badalament, R., Kane R. A. and Lee, F. (1993) *Cancer* 71, 891-898.

Nilges, M., Clore, G. M. and Gronenborn, A. M. (1988) *FEBS Lett.* 229, 317-324.

Nissen-Meyer, J. and Nes, I. F. (1997) *Arch. Microbiol.* 167, 66-77.

Noggle, J. H. and Schirmer, R. E. (1971) *The Nuclear Overhauser Effects*, Academic, New York.

Olson, A. J. and Goodsell, D. S. (1998) *Sar & Qsar in Environmental Research* 8, 273-285.

Osapay, K. and Case, D. A. (1994) *J. Biomol. NMR* 4, 215-230.

Pardi, A., Billeter, M. and Wuthrich, K. (1984) *J. Mol. Biol.* 180, 741-751.

Ploch, N. and Brawer, M. (1994) *Urology* 43, 27-35.

Quadri, L. E. N., Sailer, M., Roy, K. L., Vederas, J. C. and Stiles, M. E. (1994) *J. Biol. Chem.* 269, 12204-12211.

Quadri, L. E. N., Sailer, M., Terebiznik, M. R., Roy, K. L., Vederas, J. C. and Stiles, M. E. (1995) *J. Bacteriol.* 177, 1144-1151.

Quadri, L. E. N., Yan, L. Z., Stiles, M. E. and Vederas, J. C. (1997) *J. Biol. Chem.* 272, 3384-3388.

Rance, M., Sorensen, O., Bodenhausen, G., Wagner, G., Ernst, R. R. and Wuthrich, K. (1983) *Biochem. Biophys. Res. Commun.* 117, 479-485.

Schellhammer, P. and Wright, G. (1993) *Urol. Clinics.* 20, 597-606.

Shan, X., Gardner, K. H., Muhandiram, D. R., Rao, N. S., Arrowsmith, C. H. and Kay, L. E. (1996) *J. Am. Chem. Soc.* 118, 6570-6579.

Shawler, D. L., Bartholomew, R. M., Smith, L. M. and Dillmas. R. O. (1985*) J. Immunol.* 135, 1530-1535.

Soderberg, B. O., Sjberg, B. M., Sonnerstam, U. and Brndn, C. I. (1978) *Proc. Nat. Acad. Sci., U.S.A.* 75, 5827-5830.

Smith, L. J., Sutcliffe, M. J., Redfield, C. and Dobson, C. M. (1991) *Biochemistry* 30, 986-996.

Szyperski, T., Guntert, P., Otting, G. and Wuthrich, K. (1992) *J. Magn. Reson.* 99, 552-560.

Tichaczek, P. S., Vogel, R. F. and Hammes, W. P. (1994) *Microbiology* (UK) 140, 361-367.

Vandenbergh, P. A. (1993) *FEMS Microbiol. Rev.* 12, 221-238.

Verhoeyen, M. and Riechmann, L. (1988) *Bioessays.* 8, 74-78.

Vuister, G. W. and Bax, A. (1993) *J. Am. Chem. Soc.* 113, 7772-7777.

Wang, A. C. and Bax, A. (1995) *J. Am. Chem. Soc.* 117, 1810-1813.

Wang, A. C. and Bax, A. (1996) *J. Am. Chem. Soc.* 118, 2483-2494.

Weisemann, R., Lohr, F. and Ruterjans, H. (1994) *J. Biomol. NMR* 4, 587-593.

Wells, W. W., Pen, X. D., Yang, Y. and Rocque, P. A. (1990) *J. Biol. Chem.* 265, 15361-15364.

Williamson, M. P. and Asakura, T. (1992) *FEBS Lett.* 302, 185-188.

Williamson, M. P., Havel, T. F. and Wuthrich, K. (1985) *J. Mol. Biol.* 182, 295-315.

Winter, G. and Milstein, C. (1991) *Nature* 349, 293-299.

Wishart, D.S., Sykes, B. D. and Richards, F. M. (1991a) *J. Mol. Biol.* 222, 311-333.

Wishart, D. S. Sykes, B. D. and Richards, F. M. (1991b) *FEBS Lett.* 293, 72-80.

Wishart, D. S., Sykes, B. D. and Richards, F. M. (1992) *Biochemistry* 31, 1647-1651.

Wishart, D. S., Bigam, C. G., Holm, A., Hodges, R. S. and Sykes, B. D. (1995) *J. Biomol. NMR* 5, 67-81

Wishart, D. S. and Sykes B. D. (1994) *Methods Enzymol.* 39, 363-392.

Wishart, D. S., Bigam, C.G., Yao, J., Abildgaard, F., Dyson, H. J., Oldfield, E., Markley, J. L. and Sykes, B. D. (1995) *J. Biomol. NMR* 6, 135-140.

Wishart, D.S. and Nip, A.M. (1998) *Biochem. Cell Biol.* 76, 153-163.

Wuthrich, K. (1990) *J. Biol. Chem.* 36, 22059-22062.

Wuthrich, K. (1986) *NMR of proteins and nucleic acids*. New York, Jonh Wiley and Sons.

Yokota, T., Milenic, D. E., Whitlow, M., and Schlom, J. (1992) *Cancer Res.* 52, 3402-3408.

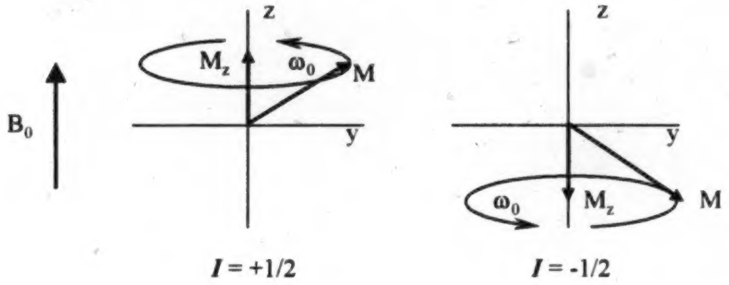

B$_0$
z
M$_z$
ω$_0$
M
y
I = +1/2
z
y
ω$_0$
M$_z$
M
I = -1/2

Figure 1.1. The two allowed spin states for a nuclear of spin I=1/2.

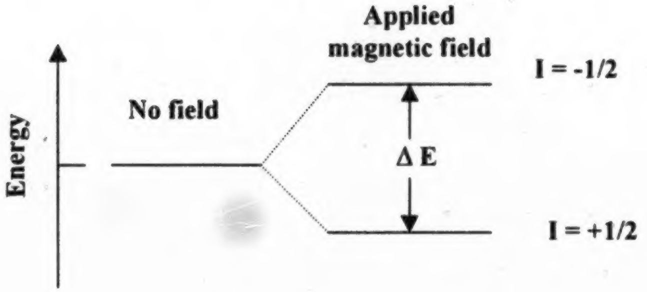

Energy
No field
Applied magnetic field
I = -1/2
Δ E
I = +1/2

Figure 1.2. Energy level diagram for spin with *I*=1/2 in a magnetic field.

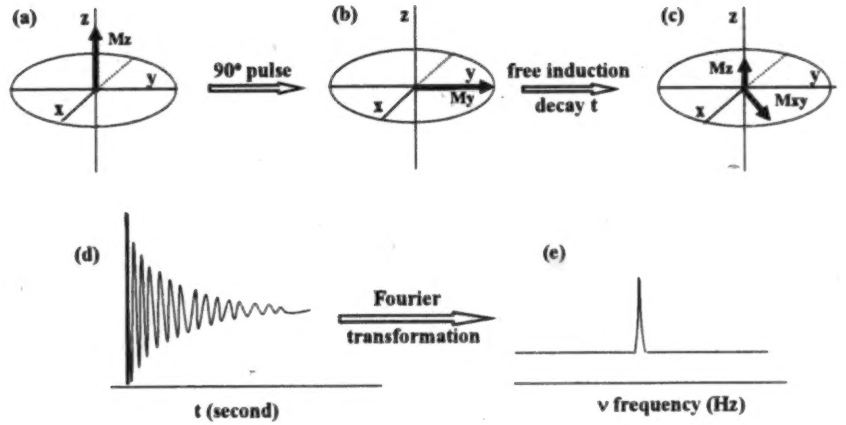

(a)
z
Mz
y
x
90° pulse
(b)
z
y
x
My
free induction
decay t
(c)
z
Mz
y
x
Mxy
(d)
t (second)
Fourier
transformation
(e)
ν frequency (Hz)

Figure 1.3. Vector picture of pulse NMR. (a) Bulk magnetization M at equilibrium in a magnetic field applied along the z-axis, (b) after a 90° pulse, the M is rotated into the xy-plane, (c) the M processes at the Larmor frequency and fades back to the equilibrium, (d) free induction decay (FID), (e) an NMR spectrum.

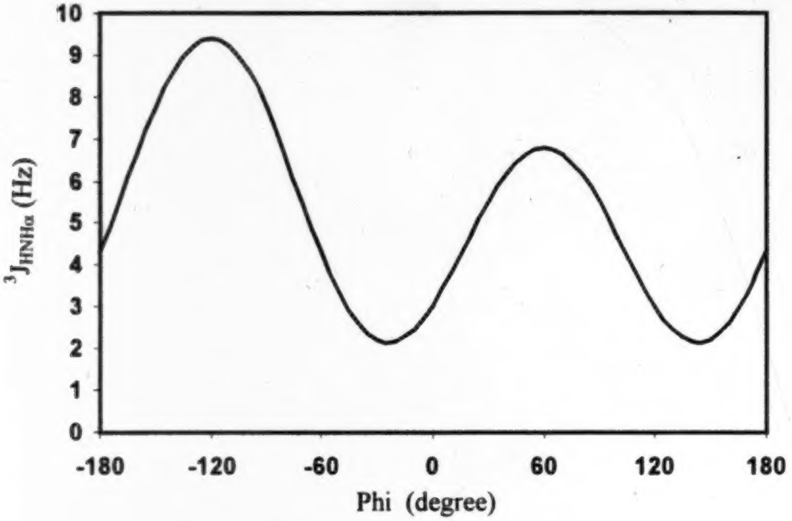


Figure 1.4. Karplus curve for $^3J_{HNH\alpha}$ as a function of the dihedral angle phi.

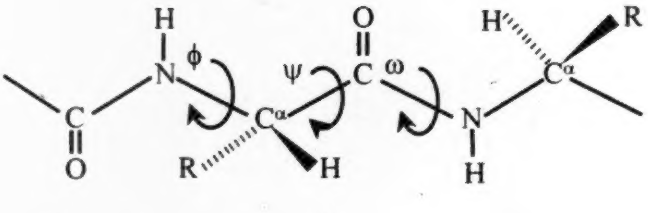

H
N
φ
C
O
ψ
Cα
R
H
O
C
ω
N
H
H
Cα
R

Figure 1.5. Standard nomenclature for the torsion angles along a polypeptide chain.

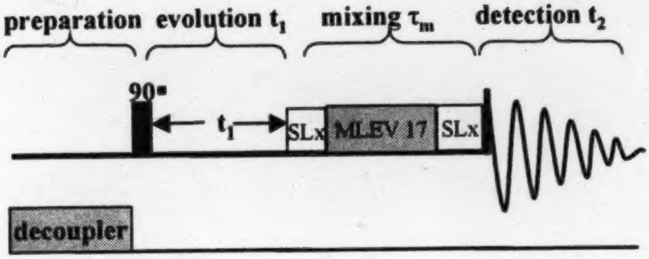

preparation
evolution t_1
mixing τ_m
detection t_2
90°
t_1
SLx
MLEV 17
SLx
decoupler

Figure 1.6. Pulse sequence for a 2D TOCSY experiment. The decoupler refers to low power irradiation at the frequency of water.

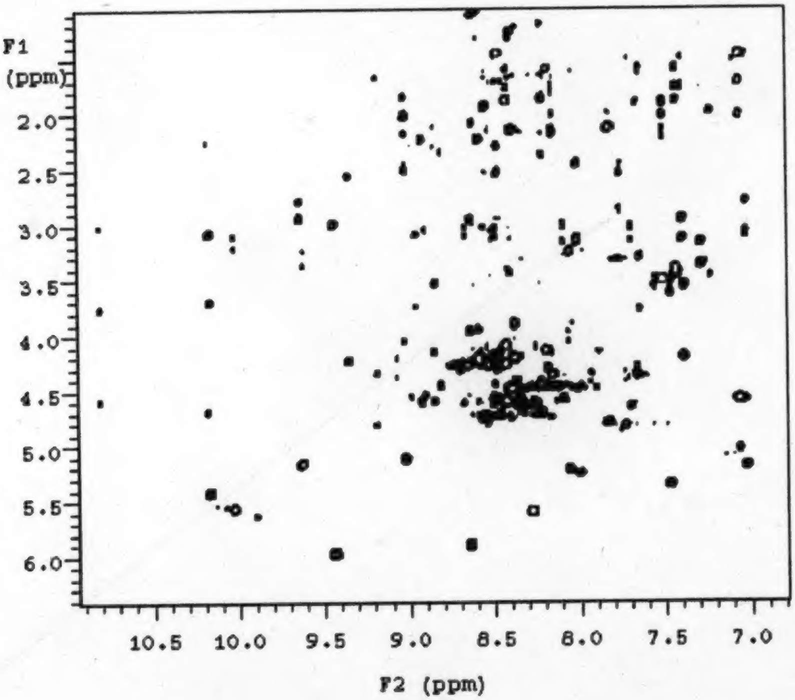


Figure 1.7. 2D TOCSY spectrum of bovine pancreatic trypsin inhibitor.

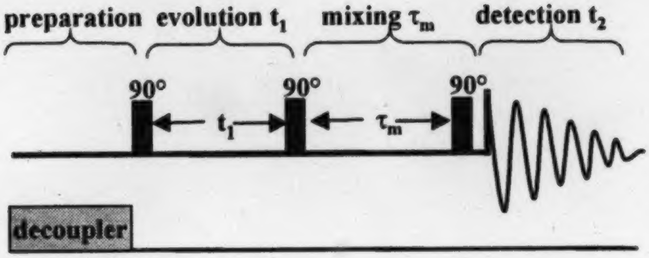

preparation
evolution t_1
mixing τ_m
detection t_2
90°
90°
90°
t_1
τ_m
decoupler

Figure 1.8. Pulse sequence for a 2D NOESY experiment. The decoupler refers to low power irradiation at the frequency of water.

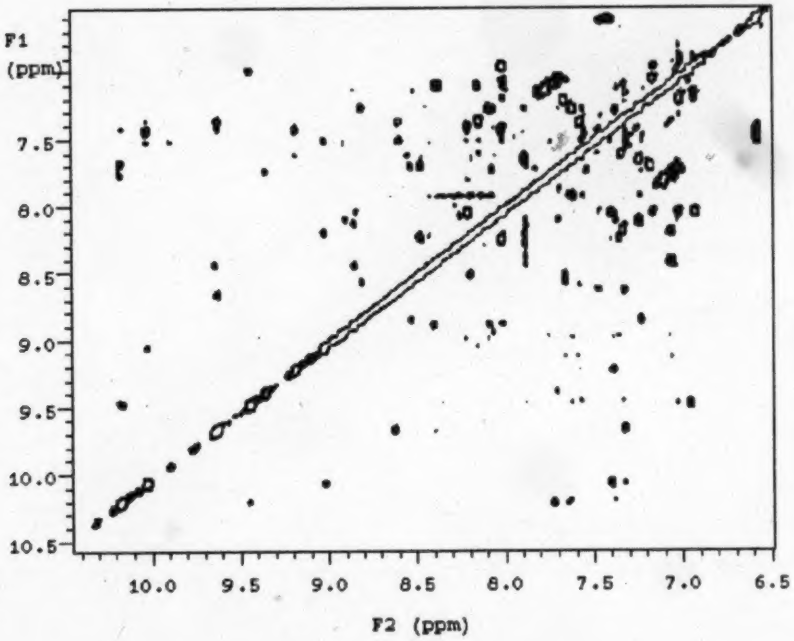


Figure 1.9. 2D NOESY spectrum of bovine pancreatic trypsin inhibitor.

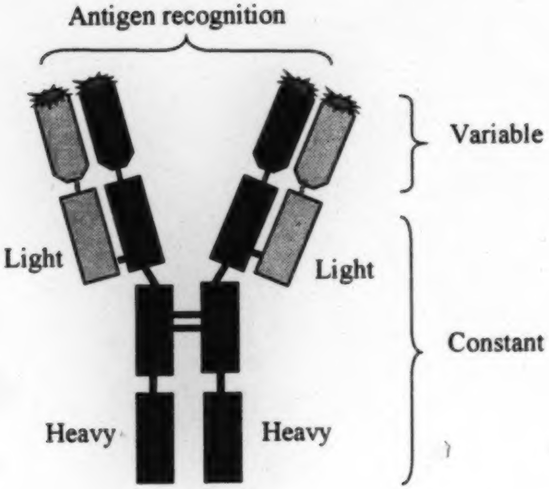

Antigen recognition
Variable
Light
Light
Constant
Heavy
Heavy

Figure 1.10. The basic unit of a natural antibody consists of two identical heavy- and two identical light-chain polypeptides. At the protein level each of these chains are folded into discrete domains.

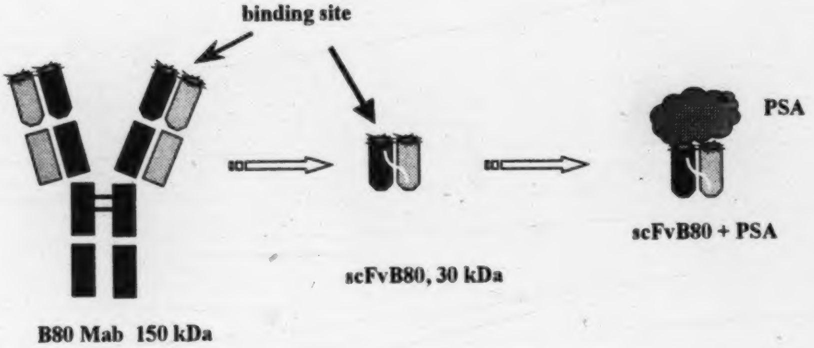

binding site
PSA
scFvB80 + PSA
scFvB80, 30 kDa
B80 Mab 150 kDa

Figure 1.11. Cartoon picture showing that scFvB80 could be used as a mimic of B80.3 Mab.

Chapter 2*

A Simple Method to Quantitatively Measure $^3J_{HNH\alpha}$ Coupling Constants from TOCSY and NOESY Spectra

2.1. Introduction

Coupling constant measurements are playing an increasingly important role in the determination and refinement of peptide and protein structures (Bax et al., 1994). Their rise in importance has had much to do with the recent introduction of innovative experimental and analytical methods which have facilitated the extraction of these otherwise difficult-to-measure parameters. A variety of ingenious curve-fitting techniques is now available for determining $^3J_{HNH\alpha}$ coupling constants from homonuclear DQF-COSY spectra (Pardi et al., 1984; Kim and Prestegard, 1989; Smith et al., 1991), from NOESY spectra (Szyperski et al., 1992) or a combination of DQF-COSY and NOESY spectra (Ludvigsen et al., 1991). Each of these methods is based on computationally intensive procedures which attempt to fit theoretical curves to anti-phase or in-phase doublets.

With the widespread use of isotopic labels in biological NMR, a number of elegant experimental methods have also been developed to measure $^3J_{HNH\alpha}$ values from ^1H-^{15}N heteronuclear experiments. These include the HMQC-J experiment (Kay and Bax, 1990), the J-modulated [^{15}N,^1H] COSY experiment (Billeter et al., 1992), the HNCA E. COSY (Weisemann et al., 1994) and the HNHA experiment (Vuister and Bax, 1993). By combining spectral data from these experiments with experiment-specific computer curve-fitting routines (Kay and Bax, 1990; Goodgame and Greer, 1993) or peak integration routines (Vuister and Bax, 1993; Billeter et al., 1992), it is possible to extract relatively accurate $^3J_{HNH\alpha}$ coupling constants. Furthermore, the use of isotopically labeled proteins, with their inherently greater chemical shift dispersion, now permits the measurement of J-coupling values of much larger proteins (up to 18 kD - Billèter et al., 1992).

Despite the arsenal of experimental and computational techniques available to quantitatively determine $^3J_{HNH\alpha}$ coupling constants, it is still quite rare to see $^3J_{HNH\alpha}$

---

*This chapter was published as a paper by Wang, Y., Nip, A.M. & Wishart, D.S. in *J. Biomol. NMR* 10, 373-382,1997.

coupling constants reported in the literature. Indeed, while there are more than 200,000 chemical shifts of peptides and proteins deposited in the BioMagResBank (Seavey et al., 1991) there are still fewer than 1300 $^3J_{HNH\alpha}$ coupling constants reported in the literature (D. Wishart, unpublished observations). The fact that quantitative (as opposed to qualitative) coupling constant measurements often require special hardware (gradients, three or four channels) or special computer programs, or the implementation of a complex pulse sequence, or the use of isotopically labeled material, or the collection and reprocessing of multiple data sets, or the complete re-assignment of "yet another spectrum" suggests that the experimental and practical barriers to measuring coupling constants are still quite significant. Even in the simplest situation (analyzing a homonuclear DQF-COSY) one is still confronted with the onerous task of re-collecting and re-assigning a spectrum which has notoriously poor signal-to-noise. Even if the signal is sufficiently good (which it rarely is), it is still necessary to use a computer program to perform a four-parameter fit over a very bumpy hyper-surface. However, an apparently successful computer fit is no guarantee of correctness, as spectral noise can often lead even the best curve-fitting program into a false local minimum.

Given these concerns and given our own frustrations at unsuccessfully implementing previously published techniques, we decided to investigate the possibility of developing a robust method for measuring $^3J_{HNH\alpha}$ coupling constants that would be: 1) simple (something that could be done in one's head or with a hand-held calculator); (2) quick (allowing $^3J_{HNH\alpha}$ determinations to be made in seconds); (3) accurate (having an RMSD of less than 1.0 Hz); (4) easy to learn; (5) independent of any requirement for specific isotopic labels; (6) applicable to conventional, easily obtained spectra (TOCSY, NOESY, COSY etc.); (7) applicable to NMR spectra with inherently high signal-to-noise (in-phase as opposed to anti-phase peaks); (8) applicable to both small peptides and large proteins; (9) generalizable to both homonuclear and heteronuclear experiments; (10) independent of any requirement for specialized hardware; (11) independent of the spectrometer make, size or type; and (12) extensively tested and verified. We believe we have developed such a method. It is based on the simple observation that $^3J_{HNH\alpha}$ coupling constants are linearly related to the linewidth at half-height ($\Delta v_{1/2}$) of in-phase TOCSY and NOESY amide crosspeaks. This method is described, in detail, below.

2.2. Materials and methods

Sample preparation and data collection. The peptides and proteins used in this study were selected on the basis of their availability, cost and the requirement that each peptide or protein had to have a high-resolution (< 2.0 Angstroms) X-ray crystal structure deposited in the Protein Data Bank (Bernstein et al., 1977). Using these criteria a total of 11 compounds was selected, including gramicidin S (five synthetic analogs), interleukin 8 (human), ubiquitin (bovine), thioredoxin (T4 phage), thioredoxin (E. coli), lysozyme (chicken) and ribonuclease A (bovine). Samples of ribonuclease A, lysozyme and ubiquitin were obtained as crystallized powders from Sigma. Samples of T4 and E. coli thioredoxin were prepared and purified as previously described (Wishart, 1991; Wishart et al., 1993). Gramicidin S analogs (small cyclic peptides with well-defined β-sheet structure) were synthesized and purified using previously published methods (Wishart et al., 1996). Five analogs were used: peptide #1 (sequence: $[PVKLF]_2$), peptide #2 (sequence: $[PVKLH)_2$), peptide #3 (sequence: $[PVKLN]_2$), peptide #4 (sequence: $[PVKLY]_2$ and peptide #5 (sequence: [GKLYPVKLYP]).

For the NMR analysis, all polypeptide samples used for this study were dissolved in 500 μL 80%H_2O/20%D_2O yielding a typical concentration of 1 - 3 mM and an uncorrected pH reading between 4.0 and 5.0. Each sample was referenced to internal DSS (Wishart et al., 1995) with the temperature being maintained at 25 ± 0.1 ^oC. The TOCSY spectra of Gramicidin S peptide #1 was collected on a Varian Unity 300 MHz spectrometer. The TOCSY and NOESY spectra of interleukin 8 were collected on a Varian Unity 600 MHz spectrometer. All other data were collected on an extensively modified Varian VXR 500 MHz spectrometer equipped with a 5 mm inverse detection probe. One-dimensional ^1H data were acquired with a ^1H sweepwidth of 6000 Hz and an acquisition time of 3.0 seconds. The residual HDO signal was suppressed by presaturation. TOCSY (Braunschweiler and Ernst, 1983; Bax and Davis, 1985) spectra were collected using 256 t_1 increments and spectral widths of 6000 Hz in both dimensions. Acquisition times were set to 0.171 seconds, relaxation delays were 2.5 seconds and spin-lock (MLEV-17) mixing times were 50 ms. The spin-lock pulse width (90^o) was 35.4 μs and the trim pulses were set to 0.5 ms. Data were zero-filled to produce a matrix of 4K x 4K complex points and processed using a shifted sine-bell weighting function (see details below) followed by base-line correction. Quadrature detection was achieved using the method of States et al. (1982). ^{1}H NOESY (Jeener et al., 1979; Kumar et al., 1980) data were collected essentially identically to the TOCSY data, with mixing times ranging from 150 to 300 ms (depending on the size of the molecule).

Assignments of all spectra were based on previously published chemical shift values (with suitable corrections for reference standards, pH and temperature). In particular, ^{1}H assignments for the gramicidin S analogs were based on those of Wishart et al. (1996), the ^{1}H assignments for E. coli thioredoxin were based on the tabulation by LeMaster and Richards (1988), the ^{1}H assignments for ubiquitin were based on those of Weber et al. (1987), the ^{1}H assignments for ribonuclease A were based on published shifts from Robertson et al. (1989), the ^{1}H assignments for hen lysozyme were based on values reported by Redfield and Dobson (1988), the ^{1}H assignments for interleukin 8 were from the report by Clore et al. (1989) and the T4 thioredoxin shifts were based on unpublished work from our laboratory (Wang et al., manuscript in preparation). On average, more than 75% of each protein's amide-alpha resonances (excluding glycines) were included in our analysis. Practical considerations of time and effort, in addition to difficulties associated with ambiguous assignments, interference from the H_2O peak and problems associated with resonance overlap generally prevented complete utilization of all resonances.

In determining the coupling constants for each polypeptide, we made use of the following Brookhaven PDB entries: E. coli thioredoxin (2TRX), ubiquitin (1UBI), ribonuclease A (8RAT), T4 thioredoxin (1AAZ), hen lysozyme (193L) and interleukin 8 (xIL8). Table 2.1 provides further details concerning the resolution and R factors for each of these crystal structures.

$^3J_{HNH\alpha}$ coupling constants for each of the above crystal structures were determined from the reported backbone ϕ angles using the following equation: $^3J_{HNH\alpha} = 5.9\cos^2\theta - 1.3\cos\theta + 2.2$, where $\theta = |\phi - 60^\circ|$. This equation, which differs slightly from previously published versions of the Karplus equation (Pardi et al., 1984; Vuister and Bax, 1993, Wang and Bax, 1996), was derived from an analysis of nearly 1000 coupling constants reported by various sources for 15 different proteins (Wishart, manuscript in preparation). Use of other widely used Karplus parameters

(Wishart, manuscript in preparation). Use of other widely used Karplus parameters led to only minor differences in the overall performance of this method. Note that in contrast to the protein samples, $^3J_{HNH\alpha}$ coupling constants for the gramicidin S analogs were obtained by direct measurement of the 1-D ^{1}H NMR spectra.

2.3. Calculation of $^3J_{HNH\alpha}$ from $\Delta\nu_{1/2}$

We have attempted to develop a protocol that should allow the determination of $^3J_{HNH\alpha}$ coupling constants independent of specific spectrometer characteristics. However, it is important to precisely follow the procedures outlined below to ensure reproducibility and accuracy:

1) After shimming the sample to get optimum lineshape and linewidth characteristics, collect a ^{1}H TOCSY or NOESY spectrum of the protein of interest. For optimal results, TOCSY spectra must be collected with an MLEV-17 mixing scheme with 0.5-2 ms trim pulses (Bax and Davis, 1985). For all spectra, acquisition times should be set such that the resolution (prior to zero-filling) in the F_2 dimension is better than 6.0 Hz/pt.

2) Assign the spectrum (using previously determined assignments or assign using conventional NOE directed techniques).

3) Zero fill in both dimensions to produce a 4K x 4K spectrum, perform a base-line correction and apply a sine-bell weighting function in the form of:

$$\sin^2[\pi(t-sbs)/2sb] \quad (2.1)$$

For Varian Spectrometers running VNMR software (Version 5.1 or higher) -

$$sb = -0.100,\ sbs = -0.066 \quad \text{(for the } F_2 \text{ dimension)}$$

Where sb and sbs are given in seconds. Processing along the F_1 dimension is somewhat less important although we generally choose values equal to 1/3 of those used for the F_2 dimension (sb1 = 0.033, sbs1 = -0.022). For other kinds of spectral processing software, the squared sine-bell weighting function (eq.2.1) can be cast into expressed as:

$$(2.2) \qquad \sin^2[\{(\pi - \phi)/T\}t + \phi]$$

where the parameters "T" and "ϕ" are related to sb and sbs as: T = 2sb + sbs and ϕ = π*sbs/2sb. In NMRPIPE (Delaglio et al., 1995), the sine-bell function is written as:

$$(2.3) \qquad \sin^2\{\pi[\mathrm{off} + \mathrm{sw}*\mathrm{t}\ (\mathrm{end} - \mathrm{off})/(\mathrm{tsize} - 1)]\}$$

where "tsize" is the total number of points in the time domain, "sw" is the sweepwidth and the parameters "off" and "end" can be expressed in terms of tsize, sw, sb and sbs as: off = -sbs/2sb and end = [(tsize - 1) - sbs*sw]/2sb*sw. Finally, for FELIX, the sinebell function is given as:

$$(2.4) \qquad \sin^2\{\pi[(1-\mathrm{wshift}/180)(\mathrm{t}*\mathrm{sw}/\mathrm{wsize}) + \mathrm{wshift}/180]\}$$

where "sw" is the sweepwidth and the phase shift "wshift" and window size "wsize" can be expressed as: wshift = -90*sbs/sb and wsize = 2sb*sw(1 - wshift/180). Where "wsize" is rounded off to the nearest integer. Using the above relationships, one can easily transform between different spectral processing systems.

4) In order to exclude the effects of passive $\alpha-\beta$ coupling constants, select F_2 traces from the upper diagonal only and determine the linewidth at half-height ($\Delta v_{1/2}$) for the central portion of each assigned ^{1}HN-^{1}Hα crosspeak (note that for NOESY spectra, any crosspeak connected to a NH resonance can be used -- not just the ^{1}HN-^{1}Hα cross peak). Consistent selection of the centre-most region of each crosspeak can greatly reduce the experimental or measurement error associated with this technique. In Varian spectrometers the command "dres" automatically determines $\Delta v_{1/2}$ for any given trace or any given 2-D peak using a simple "half-width at half-height" algorithm (least squares curve-fitting is not used). In Bruker spectrometers, $\Delta v_{1/2}$ determination takes slightly more effort.

Protocol #1

From the measured linewidth at half-height ($\Delta v_{1/2}$) substitute this value into one of the following two equations to determine the $^3J_{HNH\alpha}$ coupling constant (in Hz).

$$\textbf{(2.5)}\quad {}^3\mathbf{J}_{\mathbf{HNH}\alpha} = \mathbf{0.5(\Delta\nu_{1/2}) - MW/5{,}000 + 1.8} \quad \text{(for TOCSY data)}$$

$$\textbf{(2.6)}\quad {}^3\mathbf{J}_{\mathbf{HNH}\alpha} = \mathbf{0.6(\Delta\nu_{1/2}) - MW/5{,}000 - 0.9} \quad \text{(for NOESY data)}$$

where $\Delta\nu_{1/2}$ is the half-height linewidth in Hz of a given ^{1}HN-^{1}Hα (for TOCSY) or ^{1}HN-^{1}Hx (for NOESY) crosspeak and MW is the molecular weight of the protein in daltons. Both equations work well in practically all situations. However, care must be taken in using the correct molecular weight (i.e., is the polypeptide of interest a monomer or a dimer at NMR concentrations?) and in making sure that the temperature of the sample is between 20 °C and 35 °C. Under certain circumstances, the situation can be complicated by the presence of inherently broad linewidths, poor shimming, paramagnetic contaminants or the use of unusually high (>40 °C) or low (<20 °C) temperatures. In situations where "unusual" temperatures are used, we have found that the following linewidth correction can be employed: $\Delta\nu_{1/2} = \Delta\nu_{1/2}(\text{obs}) - 0.04(T - 25)$, where "T" is the sample temperature (in °C) and $\Delta\nu_{1/2}$(obs) is the observed linewidth (in Hz) at the given temperature. Because experimental conditions can occasionally lead to the above-mentioned complications, we elaborate on two alternative procedures that could serve as independent checks to equations (5) and (6) while at the same time eliminating the problems associated with intrinsic linewidth, temperature, sample or spectrometer differences.

Protocol #2

In situations where there is some uncertainty about the MW, it would be useful to rely on an internal linewidth standard. By identifying a resonance belonging to the protein or peptide of interest that is not affected by conformational (i.e., dihedral angle) variations, it should be possible to use its $\Delta\nu_{1/2}$ as a benchmark or reference from which to calibrate the linewidths of those resonances, such as the ^{1}HN-^{1}Hα peaks, which are affected by conformational variations. Interestingly, the doublets belonging to the 2,6 and 3,5 aromatic protons of tyrosine (~7.1 ppm) serve as ideal, easily identified, internal linewidth references. By determining the average linewidth for all assigned tyrosine crosspeaks in a TOCSY spectrum (averaging both the upper and lower diagonal signals) it is possible to use this number ($\Delta\nu_{1/2}$(Tyr)) instead of the apparent molecular weight to calculate $^3J_{HNH\alpha}$. In this case, the two best-fit equations become:

$$(2.7) \quad {}^3\mathbf{J}_{\mathbf{HNH}\alpha} = \mathbf{0.5(\Delta\nu_{1/2}) + (7.5 - \Delta\nu_{1/2}(Tyr))}$$

(for TOCSY data)

$$(2.8) \quad {}^3\mathbf{J}_{\mathbf{HNH}\alpha} = \mathbf{0.6(\Delta\nu_{1/2}) + (4.0 - \Delta\nu_{1/2}(Tyr))}$$

(for NOESY data)

where $\Delta\nu_{1/2}$ is the half-height linewidth (in Hz) of a given amide crosspeak and $\Delta\nu_{1/2}$(Tyr) is the average half-height linewidth (in Hz) for all tyrosine crosspeaks in the peptide or protein of interest (as determined from TOCSY spectra). Note that the intrinsic linewidths for tyrosine resonances in NOESY spectra tend to vary significantly; consequently TOCSY data must be used to determine $\Delta\nu_{1/2}$(Tyr).

Protocol #3

A second, albeit less refined, approach may also be used to determine the linewidth correction factor. This is based on the observation that the narrowest amide crosspeak, whether in a TOCSY or a NOESY spectrum, invariably has a ${}^3J_{HNH\alpha}$ coupling constant of close to 4.0 Hz. This phenomenon was observed for all 11 protein samples tested in this study. Furthermore, given the distribution of ϕ angles in solved protein structures from the Brookhaven Protein Databank (Bernstein et al., 1977), this observation should be generally valid for just about any structured peptide or protein one is likely to encounter (even molecules with mostly β-sheet structure). Consequently, we have found that ${}^3J_{HNH\alpha}$ can be determined quite accurately using either one of the following equations:

$$(2.9) \quad {}^3\mathbf{J}_{\mathbf{HNH}\alpha} = \mathbf{0.5(\Delta\nu_{1/2}) - 0.5(\Delta\nu_{1/2}(min)) + 4.0}$$

(for TOCSY data)

$$(2.10) \quad {}^3\mathbf{J}_{\mathbf{HNH}\alpha} = \mathbf{0.6(\Delta\nu_{1/2}) - 0.6(\Delta\nu_{1/2}(min)) + 4.0}$$

(for NOESY data)

where $\Delta\nu_{1/2}$ is the half-height linewidth (in Hz) of a given ^{1}HN-^{1}Hα (for TOCSY) or ^{1}HN-^{1}Hx (for NOESY) crosspeak and $\Delta\nu_{1/2}$(min) is the half-height linewidth (in Hz) of the narrowest amide crosspeak in the spectrum. A small disadvantage to this approach is that one cannot determine the coupling constants until after all of the resonance linewidths have been measured and the narrowest line identified. Furthermore, one must exercise caution in applying this protocol to the measurement of unstructured peptides or denatured proteins. In these situations the narrowest amide crosspeak would likely correspond to a coupling constant of 6 or 7 Hz instead of 4.0 Hz.

2.4. Results

Figure 2.1 illustrates four examples of traces taken through the F_2 dimension of ^1HN-$^1H\alpha$ and ^1HN-^1Hx crosspeaks from (a) TOCSY and (b) NOESY spectra. The $^3J_{HNH\alpha}$ value as determined from high resolution X-ray data is indicated in each figure. From these four examples it is quite clear that there is a consistent, quantitative relationship between the width of the resonance and the observed $^3J_{HNH\alpha}$ coupling constant. This can be further verified if we plot the relationship between $\Delta\nu_{1/2}$ and the $^3J_{HNH\alpha}$ coupling constant as derived from X-ray data. In Figure2.2 we show the linear relationship that exists between $\Delta\nu_{1/2}$ and $^3J_{HNH\alpha}$ for a much larger data set including all measurable resonances from (a) E. coli thioredoxin and (b) bovine ubiquitin as determined from ^1H TOCSY and ^1H NOESY spectra respectively. An excellent fit is obtained for both examples with correlation coefficients (R) of 0.94 for E. coli thioredoxin and 0.91 for ubiquitin. The strong correlation between $\Delta\nu_{1/2}$ and $^3J_{HNH\alpha}$ and the clear linear relationship observed for these and other examples suggested that a simple equation of the form:

$$^3J_{HNH\alpha} = m^*\Delta\nu_{1/2} + B \tag{2.11}$$

(where m is the slope, B is the y intercept and $\Delta\nu_{1/2}$ is the half-height linewidth) could be developed to predict coupling constants from $\Delta\nu_{1/2}$ measurements from ^1H TOCSY and NOESY spectra.

Extensive curve-fitting combined with various combinations of processing parameters allowed us to identify a common slope (m = 0.5) to all of these $^3J_{HNH\alpha}$ vs. $\Delta\nu_{1/2}$ plots for TOCSY spectra and a slope (m = 0.6) for NOESY spectra. Further comparisons revealed a clear and consistent relationship between the y-intercept (B) and the molecular weight of the peptide or protein. This relationship between the "best-fit" y-intercept (for TOCSY data) and the molecular weight of the peptide is plotted in Figure 2.3. Note that interleukin-8 forms a well-defined dimer (MW = 16.2 kD), while T4 thioredoxin (MW = 20.1 kD) shows strong evidence of forming a non-specific dimer at the concentrations used in this study. The remaining compounds are known to be monomeric. Also plotted in Figure 2.3 is the relationship between the "best-fit" y-intercept and the half-height linewidth of the aromatic protons of each compound's tyrosine resonances ($\Delta\nu_{1/2}$(Tyr)) as well as the relationship between the y-intercept and the half-height linewidth of the narrowest line ($\Delta\nu_{1/2}$(min)). Excellent fits are obtained for all three of these plots (R = 0.96, 0.91 and 0.97 respectively), indicating that equations 4 to 9 are essentially equally valid and equally accurate.

Using these equations, we were able to predict the $^3J_{HNH\alpha}$ coupling constant for a total of 11 peptides and proteins. The correlation between these predicted coupling constants (designated as J_{lw} - since they were derived from linewidth measurements) and the coupling constants derived from the corresponding high resolution X-ray structures (designated as J_{xray}) are shown in Tables 2.2 and 2.3. These two tables summarize the results of more than 650 coupling constant measurements (383 from TOCSY data; 279 from NOESY data) made using this simple technique. Both tables clearly show the excellent agreement obtained for both large (20 kD) and small (1 kD) polypeptides and that this agreement is consistently good from experiment to experiment and protein to protein. Overall, for the 11 polypeptides tested, TOCSY data yielded an average correlation coefficient of 0.89 and an RMSD from J_{xray} of 0.86 Hz while NOESY data yielded an average correlation coefficient of 0.90 and an RMSD from J_{xray} of 0.85 Hz.

2.5. Discussion

The methods described in this manuscript offer a unique approach to quantitatively measuring J-coupling constants. Almost every method previously published depends on the measurement of peak-to-peak separation (Pardi et al., 1984; Kim and Prestegard, 1989; Ludvigsen et al., 1991; Szyperski et al., 1992) combined with detailed computer-aided curve-fitting to determine J-coupling constants. In our approach, no attempt is made to measure peak-to-peak separations. Rather, the half-height linewidth ($\Delta\nu_{1/2}$) of in-phase peaks is used as a simple proxy for peak-to-peak separation. This avoids the difficulties and frailties of earlier methods because linewidths are less affected by noise, peak intensity and digital resolution than are peak positions. This is underlined by the fact that our spectra could be analyzed at a digital resolution (after zero filling) of 3.0 Hz/pt. while most DQF-COSY analysis, for instance, requires a digital resolution of 0.5 Hz/pt. or better.

While a relationship between $\Delta\nu_{1/2}$ and $^3J_{HNH\alpha}$ was not unexpected, what was particularly surprising was that this relationship was so linear and that the same general equation ($J = mx + B$) could be applied to such a range of differently sized molecules. Simulations, using the same weighting functions as employed in this paper, on spectral doublets having a variety of linewidths, revealed that the relationship between $\Delta\nu_{1/2}$ and $^3J_{HNH\alpha}$ is only approximately linear (see Figure 2.4). Furthermore, these simulations predict that the slope of the curve (a parabola) should increase with increasing molecular weight (or decreasing T_2). Given the gradual nature of the predicted curve, we believe the scatter associated with our plots is probably too great to distinguish between a straight line and a parabola. Consequently a linear approximation, as we have employed in this paper, is entirely adequate to quantitatively predict J-coupling constants. Interestingly, we were unable to detect a trend in our experimental data which reiterated the predicted relationship between the slope and the molecular weight. This difference between experimental and theoretical results may have to await further investigation. Nevertheless, the apparent independence between the slope and molecular weight (or T_2) certainly makes protocols we have described much easier to use and far easier to remember.

Because this method makes use of linewidth measurements, and because linewidths are sensitive to T_2's, correlation times, temperature, molecular weights, sample conditions and shimming, it is important that appropriate correction factors be identified and properly used. We have described three independent methods for determining these correction factors. One is based on molecular weight (protocol #1), another is based on an internal linewidth standard (protocol #2) and the third is based on the linewidth of the narrowest amide crosspeak (protocol #3). Protocol #1 is perhaps the simplest to use and is our preferred choice in deriving coupling constants from TOCSY or NOESY data. Protocol #2 is perhaps the most rigorous and least sensitive to sample or spectrometer variations. Its only disadvantage is that it requires a few additional linewidth measurements for calibration purposes. We recommend that protocol #3 be used only if there are no tyrosines in the molecule or if one wishes to perform an independent check of protocols #1 or #2. The option of choosing any one of three independent methods for calculating $^3J_{HNH\alpha}$ coupling constants certainly provides a level of redundancy and robustness not often found in many other NMR measurements.

In assessing the accuracy of this method we were careful to select a subset of polypeptides spanning a significant size range (1 kD to 20 kD) for which there were readily available high resolution X-ray structures. It has been noted by others (Garrett et al., 1994; Smith et al., 1991; Pardi et al., 1984) that higher resolution X-ray structures invariably yield better agreement with measured J-coupling constants than lower resolution X-ray structures. Furthermore, X-ray structures (no matter what resolution) consistently show better agreement between measured J-coupling constants than structures generated through NMR distance geometry methods (Garrett et al., 1994). Indeed, our own investigations on almost 1000 $^3J_{HNH\alpha}$ coupling constants extracted from the literature (D.S. Wishart, manuscript in preparation) show that the standard deviation between measured $^3J_{HNH\alpha}$ values and those derived from NMR-generated structures is typically greater than 1.69 Hz, while the standard deviation between measured $^3J_{HNH\alpha}$ values and those derived from X-ray structures is less than 0.96 Hz. This difference is quite substantial and certainly justifies our reliance on X-ray dihedral angles rather than NMR-derived dihedral angles to assess the accuracy of our method.

Nevertheless, differences between one X-ray crystal form and another (say tetragonal versus triclinic) or between one structure at high resolution and another at marginally lower resolution, typically introduce a mean error (averaged over the length of the polypeptide) of $\pm 10^{\circ}$ for the ϕ dihedral angle. This suggests that the best possible agreement between crystal-structure derived J-coupling constants and a "perfect" method for determining coupling constants directly from NMR data would lead to a "best-case" RMSD of 0.54 Hz and a correlation coefficient of slightly less than 0.98 (Wang and Bax, 1996). A more reasonable set of values, which would account for the inherent differences between solution and crystal structures, would likely add a further 0.50 Hz to the uncertainty in $^3J_{HNH\alpha}$, thereby giving a predicted correlation coefficient of 0.93. With these theoretical limits in mind, we compared our results to both these predictions and to the results from other popular methods for quantitatively measuring $^3J_{HNH\alpha}$ coupling constants.

In total, more than 650 $^3J_{HNH\alpha}$ coupling constants were evaluated from our data set. This is roughly equal to 1/2 of all (quantitative) $^3J_{HNH\alpha}$ coupling constants reported over the past 20 years. On average, our experimentally measured J_{lw} values agreed with derived J_{xray} values with a correlation coefficient (R) of 0.89 and a RMSD error of 0.86 Hz. This compares very favorably to the results obtained by Kim and Prestegard (1989) for their DQF-COSY method where R=0.74 and RMSD = 1.95 Hz for acyl carrier protein. It also compares favorably with the results reported by Smith et al. (1991) wherein their method yielded R=0.96 and RMSD = 0.93 Hz for DQF-COSY data collected on hen lysozyme. Similarly, Ludvigsen et al. (1991) obtained an R value of 0.90 and an RMSD = 1.16 Hz for their NOESY/DQ-COSY method as applied to the CI-2 inhibitor. The HMQC-J approach (Kay et al., 1989) when applied to Staphylococcal nuclease yielded an R value of 0.89 and an RMSD = 1.01 Hz. The HNHA method of Vuister and Bax (1993) as similarly applied to Staphylococcal nuclease yielded an R value of 0.91 and an RMSD = 0.76 Hz.

With the possible exception of the method of Kim and Prestegard (1989), nearly all of the methods (including ours) achieve a level of agreement that is reasonably close to optimal (RMSD = 0.54 Hz, R = 0.96) and essentially identical to expected (RMSD = 1.04 Hz, R = 0.93). Overall, the average correlation coefficient for the four best methods was R = 0.92 and the RMS deviation was 0.97 Hz. While the average correlation coefficient for the four best methods is slightly higher than ours (0.92 vs. 0.89), it is important to note that our calculations were performed on a substantially larger (five to 15 times larger) sample and a significantly more diverse set of polypeptides (in both size and structure) than any of the other methods. Had any of these previously published methods been applied to a comparable number of data points (>650) or to proteins as large (>20 kD) as those analyzed here, we expect the results would have been somewhat different. In this regard, the work of Garrett et al. (1994) is of particular note. These workers recently re-evaluated the HNHA experiment using a much larger data set (264 points vs. 96 points) and found that the RMS deviation for this more representative sample increased to 1.42 Hz (from 0.76

Hz) and the correlation coefficient fell from 0.91 to 0.85.

To summarize, we have described a novel method that allows $^3J_{HNH\alpha}$ coupling constants to be rapidly determined from simple linewidth measurements of ^1H TOCSY or ^1H NOESY spectra. This method makes use of the linear relationship between $^3J_{HNH\alpha}$ and $\Delta\nu_{1/2}$ of appropriately processed NMR spectra. We believe this new approach offers several significant advantages over most other published techniques.

First, it is simple enough that coupling constants can be determined almost by inspection and without the need of a computer program or any kind of complex curve-fitting routine. Indeed, only a quick calculation based on the measured linewidth of amide ^1H traces is needed to obtain an accurate $^3J_{HNH\alpha}$ coupling constant.

Second, the method is very quick. Once the assignment process has been completed we have generally found it to be possible to manually determine the $^3J_{HNH\alpha}$ coupling constants of a 100 residue protein in less than 30 minutes. This has allowed one of us (Y.W.) to measure more than 650 $^3J_{HNH\alpha}$ coupling constants -- nearly 1/2 of the total of all $^3J_{HNH\alpha}$ values ever reported over the last 20 years -- in less than a few hours (this total, however, does not include the considerable time required to prepare the samples, collect the spectra and assign all of the proteins used in this study).

Third, the method is accurate. As described previously, this new approach has an RMSD of less than 0.9 Hz and a correlation coefficient of 0.89 when compared to X-ray-derived coupling constants. This equals or betters the performance or accuracy of nearly every other method published to date.

Fourth, the method is applicable to conventional, easily obtained, high signal-to-noise NMR experiments. In particular, we have shown that this method works well for both TOCSY and NOESY data. These are robust NMR experiments which can be routinely collected on almost any modern high-field spectrometer. Further, TOCSY and NOESY spectra have signal-to-noise ratios that are minimally 16 times higher than conventional DQF-COSY spectra, thereby allowing higher quality J-coupling data to be collected at a much faster rate than DQF-COSY's.

Fifth, the method is applicable to both small peptides and large proteins. In particular, we have been able to obtain accurate coupling constants for proteins as large as 20.1 kD (T4 thioredoxin) using only homonuclear TOCSY data. Previously, the largest protein or protein complex for which quantitative coupling constants have been reported was 18 kD (Billeter et al., 1992) and this complex required the use of heteronuclear spectroscopy. We have every reason to believe that our technique could be applied to molecules significantly larger than 20 kD.

Sixth, the method is independent of special requirements in terms of isotopic labeling (we are currently developing approaches to apply this method to ^{15}N labeled material), spectrometer hardware (third and fourth channels are not needed, nor are pulsed-field gradients) and spectrometer software (special pulse sequences and special deconvolution programs are not required). This generalizability makes this method highly portable and easily implemented -- even by the most naive NMR spectroscopist.

Seventh, the method has been thoroughly tested. Having measured and verified more than 650 coupling constants from 11 different polypeptides (from 1 kD to 20 kD) collected at three different field strengths (300 MHz, 500 MHz and 600 MHz) over a range of temperatures (20 °C to 35 °C), we believe this approach to coupling constant measurement is among the most thoroughly tested methods yet presented.

While there are many positive aspects to this simple approach to $^3J_{HNH\alpha}$ coupling constant determination, it is still important to remember that it is not without some flaws. Certainly there may arise circumstances where protein linewidths, for whatever reason, may become sufficiently large (perhaps > 30 Hz) such that the general linear relationship (Eq. 11) does not hold. Similarly, for very small peptides (< 10 residues), it is often possible to measure the peak-to-peak separation of in-phase TOCSY or NOESY doublets without having to measure their linewidths. In these extreme situations it is neither practical nor particularly useful to apply linewidth analysis. It is also important to note that even under ideal circumstances, it is possible to introduce a systematic error (up to 0.5 Hz) in coupling constant measurements through an incorrect determination of the "y-intercept" or correction factor. Care, therefore, must be taken to ensure that this correction factor is consistent with what is known about the molecule (i.e. is it a dimer or a monomer?) and that it yields a range of $^3J_{HNH\alpha}$ values typical for proteins (between 3 Hz and 10 Hz).

Still another limitation with this approach to $^3J_{HNH\alpha}$ measurement is the fact that it requires TOCSY or NOESY spectra to be collected and processed in a very precise manner. Obviously computer-based curve-fitting programs are more flexible and do not typically constrain the user to follow special spectral collection and processing conditions. Nevertheless, we have found that our approach is somewhat more flexible than what might be expected. In this regard, we investigated whether measured linewidths were sensitive to different levels of digital resolution (both before and after zero filling). Using three sets of ubiquitin TOCSY data with digital resolutions ranging from 6 Hz/pt to 1.5 Hz/pt (before and after zero filling) we found that measured linewidths for all three spectra were essentially identical with an average RMS difference of less than 0.2 Hz and a correlation coefficient of 0.997. Consequently, the coupling constants predicted from these three data sets were essentially identical. This indicates that spectral resolution (so long as the intrinsic linewidth is not a result of poor digitization) does not seem to adversely affect the lineshapes, the linewidths or the predicted coupling constants. In addition to this work on digital resolution, we investigated how the use of different kinds of signal processing functions (gaussian, shifted gaussian, shifted sinebell, etc.) on TOCSY and NOESY data might affect the correlation between $\Delta\nu_{1/2}$ and $^3J_{HNH\alpha}$. While the numbers for the slope and intercept do change, the linear relationship between linewidth and $^3J_{HNH\alpha}$ appears to hold for all processing parameters so far tested.

Because the method we have described is fundamentally based on linewidth measurements and because linewidths are sensitive to segmental motions, fast and slow exchange, polymer "end" effects, spectral overlap and decoupler distortion, it is important to be aware of the complications that these phenomena can cause. For instance, if the protein has a very flexible head or tail, or if it contains a mobile "hinge" region, then it is likely that the linewidths for these segments will be somewhat different from the rest of the protein. Because the peptides and proteins we selected were all stable, well-structured, single domain molecules, we did not encounter this problem. However, we did observe amide resonances at the C-termini of two proteins that were somewhat narrower than other resonances. Similarly we found at least one occurrence where the linewidth of a particular resonance was nearly twice as broad (40 Hz!) as the next widest resonance. Whether this was due to exchange broadening, decoupler distortion or spectral overlap is not clear, but the presence of such an outlier is usually sufficiently obvious that it can be dealt with appropriately.

One unexpected result from this work was the observation that the linewidths from TOCSY crosspeaks are generally narrower than NOESY crosspeaks. This observation led to the development of separate equations relating $^3J_{HNH\alpha}$ coupling constants to measured TOCSY and NOESY linewidths. To understand why this difference was observed, it is important to remember that TOCSY crosspeaks, unlike NOESY crosspeaks, have a mixture of both in-phase and anti-phase components (Bax and Davis, 1985). These anti-phase components, if not completely removed, will cause slight phase distortions in TOCSY line shapes (this distortion is most obvious for glycines) which will lead to changes (i.e. narrowing) in the apparent linewidth of an unresolved doublet. The only way to remove these anti-phase components is to employ z-filtering (Sorensen et al., 1984). As shown by Subramanian and Bax (1987) the use of z-filters in a 1-D TOCSY experiment can yield spectra that are sufficiently free from distortion to permit very accurate peak-to-peak measurements of well-resolved multiplets. To investigate this issue further, z-filtered TOCSY spectra employing a DIPSI mixing sequence were collected for a mid-sized protein (ubiquitin) and compared to both "unfiltered" TOCSY and conventional NOESY spectra of the same molecule. Results indicate that the use of a z-filtered TOCSY experiment led to spectra with linewidths very similar to those measured from our NOESY spectra. This suggests that if one were to employ z-filtered TOCSY experiments (instead of the unfiltered ones employed in this study) that, in all likelihood, only a single set of equations would be needed to extract coupling constants from either NOESY or TOCSY spectra. This result underlines the dependence that this particular technique has on the spectral collection conditions and serves to emphasize the importance of adhering to the collection and processing conditions described in the Methods section of this manuscript.

2.6. Conclusion

As this simple method of $^3J_{HNH\alpha}$ coupling constant determination is applied to other biomolecular systems, we expect further refinements and improvements will be possible. We are also hopeful that this very general concept of linewidth measurement, as opposed to peak-to-peak measurement, will find applications beyond the determination of protein $^3J_{HNH\alpha}$ coupling constants through homonuclear ^1H spectroscopy. Currently we are working on methods to quantitatively measure $^3J_{HNH\alpha}$ coupling constants using heteronuclear techniques. We are also working to develop methods for measuring other vicinal and geminal coupling constants ($J_{\alpha\beta}$) using a combination of both homonuclear and heteronuclear spectroscopy. We believe that this and other related work will make quantitative coupling constant measurements far simpler and far easier to use in analyzing the conformation of peptides, proteins, carbohydrates and other biomolecules through NMR.

References

Baldwin, E. T., Weber, I. T., St. Charles, R., Xuan, J. C., Appella, E., Yamada, K., Matsushima, K., Edwards, B. F., Clore, G. M., Gronenborn, A. M. and Wlodawer, A. (1991) *Proc. Natl. Acad. Sci.* 88, 502-506.

Bax, A. and Davis, D. G (1985) *J. Magn. Reson.* 65, 355-360.

Bax, A., Vuister, G. W., Grzesiek, S., Delaglio, F., Wang, A. C., Tschudin, R. and Zhu, G. (1994) *Meth. Enzymol.* 239, 79-105.

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Brice, M. D., Rogers, J. R., Kennard, O., Shimanouchi, T. and Tasumi, M. J. (1977) *J. Mol. Biol.* 112, 535-542.

Billeter, N., Neri, D., Otting, G., Qian, Y. Q. and Wuthrich, K. (1992) *J. Biomol. NMR* 2, 257-274.

Braunschweiler, L. and Ernst, R. R. (1983) *J. Magn. Reson.* 53, 521-528.

Clore, G. M., Appella, E., Yamada, M., Matsushima, K. and Gronenborn, A. M. (1989) *J. Biol. Chem.* 264, 18907-18911.

Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. and Bax, A. (1995) *J. Biomol. NMR* 6, 277-293.

Garrett, D. S., Kuszewski, J., Hancock, T. J., Lodi, P.T., Vuister, G. W., Gronenborn, A. M. and Clore, G. M. (1994) *J. Magn. Reson. B.* 104, 99-103.

Goodgame, M. M. and Greer, S. M. (1993) *J. Magn. Reson. A.* 102, 246-248.

Jeener, J., Meier, B. H., Bachmann, P. and Ernst, R. R. (1979) *J. Chem. Phys.* 71, 4546-4553.

Katti, S. K., LeMaster, D. M. and Eklund, H. (1990) *J. Mol. Biol.* 212, 167-184.

Kay, L. E., Brooks, B., Sparks, S. W., Torchia, D. A. and Bax, A. (1989) *J. Am. Chem. Soc.* 111, 5488-5490.

Kay, L. E. and Bax, A. (1990) *J. Magn. Reson.* 86, 110-126.

Kim, Y. and Prestegard, J. H. (1989) *J. Magn. Reson.* 84, 9-13.

Kumar, A. Ernst, R. R. and Wuthrich, K. (1980) *Biochem. Biophys. Res. Commun.* 95, 1-6.

LeMaster, D. M. and Richards, F. M. (1988) *Biochemistry* 27, 142-150.

Ludvigsen, S., Andersen, K. V. and Poulsen, F. M. (1991) *J. Mol. Biol.* 217, 731-

736.

Pardi, A., Billeter, M. and Wuthrich, K. (1984) *J. Mol. Biol.* 180, 741-751.

Redfield, C. and Dobson, C. M. (1988) *Biochemistry* 27, 122-136.

Robertson, A. D., Purisima, E. O., Eastman, M. A. and Scheraga, H. A. (1989) *Biochemistry* 28, 5930-5938.

Seavey, B. R., Farr, E. A., Westler, W. M. and Markely, J. L. (1991) *J. Biomol. NMR* 1, 217-236.

Smith, L. J., Sutcliffe, M. J., Redfield, S. C. and Dobson, C. M. (1991) *Biochemistry* 30, 986-996.

Sorensen, O. W., Rance, M. and Ernst, R. R. (1984) *J. Magn. Reson.* 56, 527-533.

States, D. J., Haberkorn, R. A. and Ruben, D. J. (1982) *J. Magn. Reson.* 48, 286-292.

Subramanian, S. and Bax, A. (1987) *J. Magn. Reson.* 71, 325-330.

Szyperski, T., Guntert, P., Otting, G. and Wuthrich, K. (1992) *J. Magn. Reson.* 99, 552-560.

Tilton, R. F., Dewan, J. C. and Petsko, G. A. (1992) *Biochemistry* 31, 2469-2481.

Vijay-Kumar, S., Bugg, C. E. and Cook W. J. (1987) *J. Mol. Biol.* 194, 531-544.

Vuister, G.W. and Bax, A. (1993) *J. Am. Chem. Soc.* 113, 7772-7777.

Wang, A.C. and Bax, A. (1996) *J. Am. Chem. Soc.* 118, 2483-2494.

Weber, P. L., Brown, S. C. and Mueller, L. (1987) *Biochemistry* 26, 7282-7290.

Weisemann, R., Ruterjans, J., Schalbe, H., Schleucher, J., Bermel, W. and Griesinger, C. (1994) *J. Biomol. NMR* 4, 231-240.

Wishart, D. S. (1991) Ph.D. Thesis, Yale University, New Haven.

Wishart, D. S., Bigam, C. G., Yao, J., Abildgaard, F., Dyson, H. J., Oldfield, E., Markley, J. L. and Sykes, B. D. (1995) *J. Biomol. NMR* 6, 135-140.

Wishart, D. S., Kondejewski, L. H., Semchuk, P. D., Sykes, B. D. and Hodges, R.S. (1996) *Lett. Pept. Science* 3, 53-60.

Wishart, D. S., Sykes, B. D. and Richards F. M. (1993) *Biochim. Biophys. Acta* 1164, 36-46.

Young, A. C. M., Tilton, R. F. and Dewan, J. C. (1994) *J. Mol. Biol.* 235, 302-317.

Table 2.1. Listing of high resolution X-ray structures used in calculating J_{xray} values.

Protein	Accession	Resolution	R-factor	Reference
Thioredoxin (E. coli)	2TRX	1.68 Å	0.165	Katti, S.K. et al. (1991)
Ubiquitin (Bovine)	1UBI	1.80 Å	0.165	Vijay-Kumar, S. et al. (1987)
Ribonuclease A (Bovine)	8RAT	1.50 Å	0.158	Tilton, R.F. et al. (1992)
Thioredoxin (T4 phage)	1AAZ	2.00 Å	0.210	Ekland, H. et al. (unpublished)
Lysozyme (Chicken)	193L	1.33 Å	0.184	Young, A.C.M. et al. (1994)
Interleukin 8 (Human)	xIL8	1.50 Å	0.180	Baldwin, E.T. et al. (1991)

Table 2.2 Summary of results obtained using linewidth analysis (Protocol #1) for determination of ${}^3J_{HNH\alpha}$ coupling constants (TOCSY data only).

Protein	MW (daltons)	No. of points	R (J_{xray} vs. J_{lw})	RMSD (J_{xray} vs. J_{lw})
T4 Thioredoxin	20,100	47	0.88	0.88 Hz
Interleukin 8	16,200	47	0.91	0.88 Hz
Lysozyme	14,300	87	0.85	0.94 Hz
Ribonuclease A	13,700	69	0.86	0.98 Hz
E.C. Thioredoxin	11,700	62	0.94	0.72 Hz
Ubiquitin	8,600	48	0.91	0.81 Hz
GS peptide #1-#4	1200	16	0.96	0.74 Hz
GS peptide #5	1200	7	0.96	0.50 Hz

Table 2.3. Summary of results obtained using linewidth analysis (Protocol #1) for determination of ${}^3J_{HNH\alpha}$ coupling constants (NOESY data only).

Protein	MW (daltons)	No. of points	R (J_{xray} vs. J_{lw})	RMSD (J_{xray} vs. J_{lw})
Interleukin 8	16,200	46	0.89	0.93 Hz
Lysozyme	14,300	80	0.88	0.86 Hz
Ribonuclease A	13,700	46	0.92	0.85 Hz
E.C. Thioredoxin	11,700	53	0.91	0.83 Hz
Ubiquitin	8,600	54	0.91	0.85 Hz

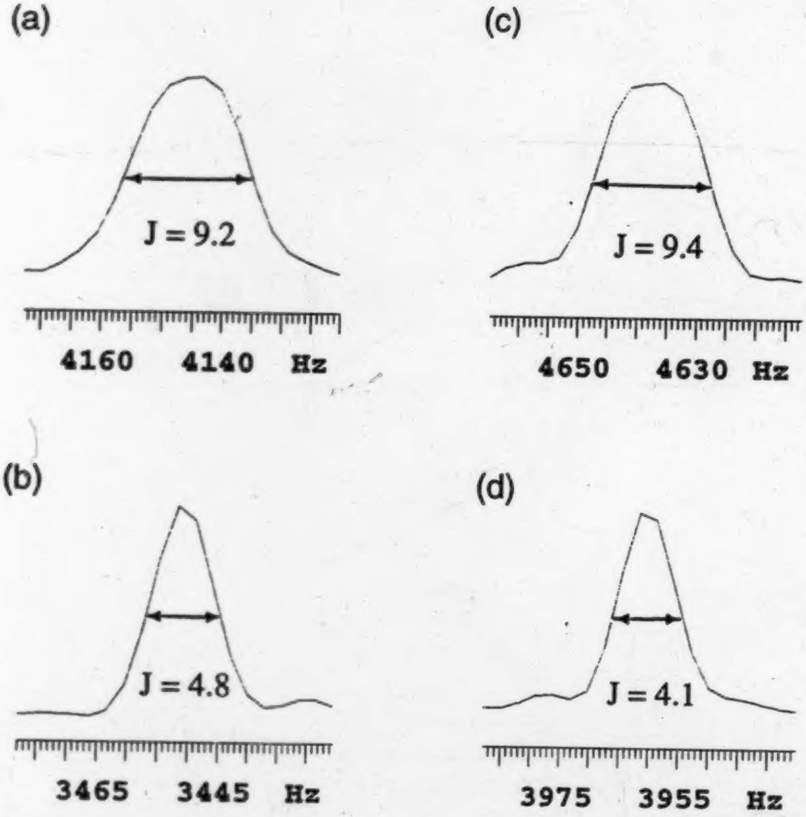


Figure 2.1. Four examples of traces taken through the F_2 dimension of ^{1}HN-^{1}Hα and ^{1}HN-^{1}Hx crosspeaks from TOCSY and NOESY spectra. Illustrated in a) and (b) are TOCSY traces of amide crosspeaks from Ribonuclease A (Asn^{103}: $\Delta\nu_{1/2} = 21.1$ Hz; Ser^{90}: $\Delta\nu_{1/2} = 11.7$ Hz). Illustrated in (c) and (d) are NOESY traces of amide crosspeaks from Ubiquitin (Val^5: $\Delta\nu_{1/2} = 19.9$ Hz; His^{58}: $\Delta\nu_{1/2} = 11.2$ Hz). The $^3J_{HNH\alpha}$ value (in Hz) as determined from high resolution X-ray data is indicated in each figure. Note that broad peaks are associated with large coupling constants while narrow peaks are associated with small coupling constants.

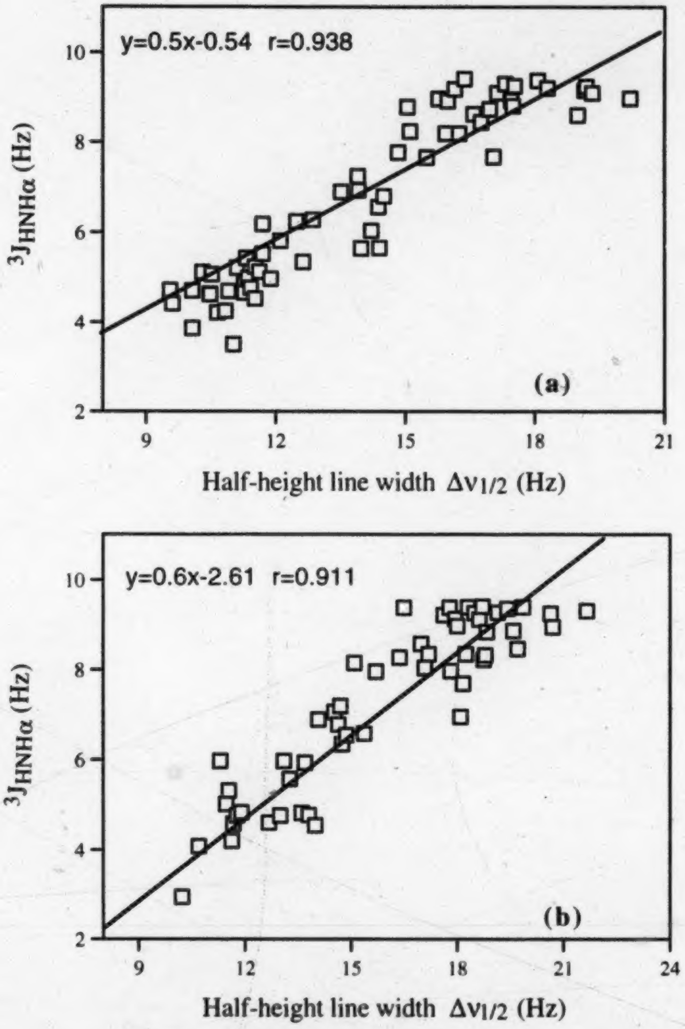


Figure 2.2. Relationship between $\Delta\nu_{1/2}$ and $^3J_{HNH\alpha}$ for (a) E. coli Thioredoxin and (b) Bovine Ubiquitin as determined from ^{1}H TOCSY and ^{1}H NOESY spectra respectively. The equation for the "best-fit" line derived from Protocol #1 and the correlation coefficient (r) are shown in the top left corner of each graph. Note that the superimposed line is a "best fit" line for all of the data (650 points) and all of the proteins (11) and so, for any given protein, there may be slight systematic deviations at certain extrema.

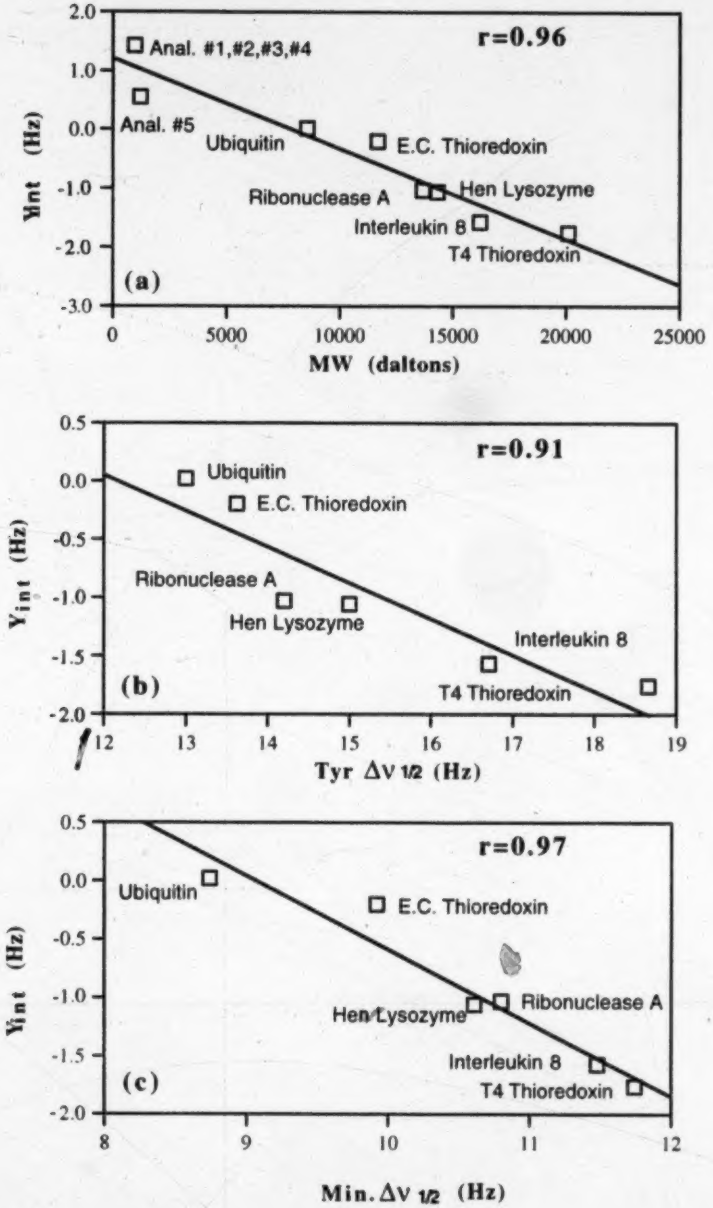


Figure 2.3. Relationship between the "best-fit" y-intercept (for TOCSY data) and (a) the molecular weight of the peptide; (b) the tyrosine half-height linewidth; and (c) the linewidth of the narrowest amide resonance. The correlation coefficient (r) for each line is given in the top right corner of each graph.

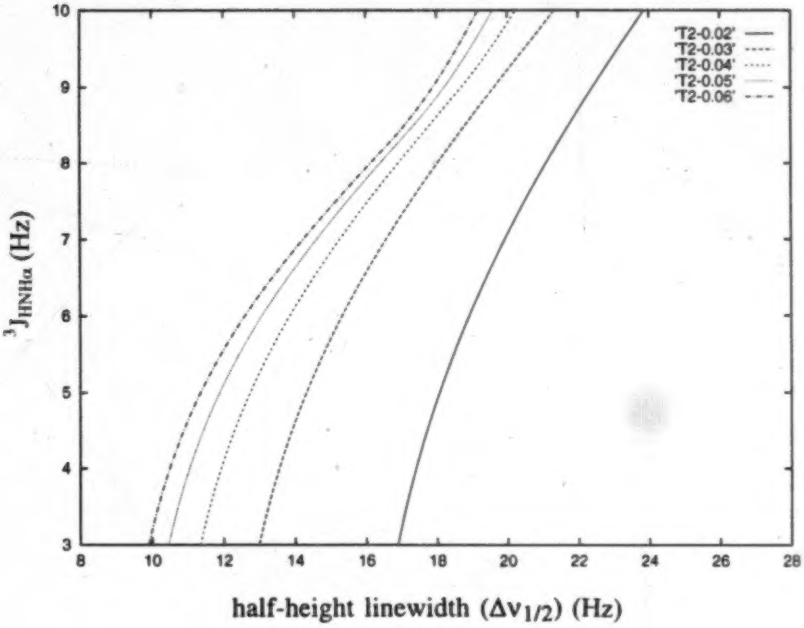


Figure 2.4. Computer simulation of the dependence of the ${}^3J_{HNH\alpha}$ coupling constant on the half-height linewidth ($\Delta\nu_{1/2}$) using six different T_2 values.

Chapter 3*

Facile Measurement of Polypeptide $^3J_{HNH\alpha}$ Coupling Constants from HMQC-J Spectra

3.1. Introduction

Since its introduction in 1989 (Kay et al., 1989; Kay and Bax, 1990) the HMQC-J experiment has emerged as one of the most popular methods for extracting precise $^3J_{HNH\alpha}$ coupling constants from polypeptide spectra. By taking advantage of the longer T_2 relaxation time found for ^{15}N nuclei, Kay and co-workers were able to show that a highly digitized HMQC experiment could provide sufficient resolution to accurately measure $^3J_{HNH\alpha}$ coupling constants. This remarkably simple approach, which exploits two key advantages of ^{15}N nuclei (wide chemical shift dispersion and long T_2's), has inspired the development of a host of other heteronuclear experiments for J-coupling measurement (Billeter et al., 1992; Vuister and Bax, 1993; Weisemann et al., 1994).

While the HMQC-J experiment is relatively trivial to implement, the extraction of coupling constants from HMQC-J data is not quite so simple. The strong resolution enhancement required to differentiate in-phase doublets introduces two major problems: (i) reduced signal intensity and (ii) non-linearity in the relationship between $^3J_{HNH\alpha}$ and the peak-to-peak separation. While little can be done to address the issue of reduced signal-to-noise, three different computational methods have been proposed to deal with problem of non-linearity (Kay and Bax, 1990; Forman-Kay et al., 1990; Goodgame and Geer, 1993). Essentially all three approaches require the spectroscopist to process HMQC-J spectra using a progressive array of line-narrowing (resolution enhancement) filters. By measuring how the peak-to-peak separation changes as a function of the line-narrowing filters and then fitting these results to a simulated curve, it is possible to determine $^3J_{HNH\alpha}$ values to relatively good accuracy. However, this iterative fitting process can be both tedious and error-prone. For instance, with the procedure of Goodgame and Geer, a 150-residue protein would require measuring and recording the peak-to-peak separation of approximately 900 (6 x 150) doublets. In our hands, this protocol often requires a very full day of

---

*This chapter was published as a paper by Wishart, D.S. & Wang, Y. in *J. Biomol. NMR* 11, 329-336,1998.

intensive data processing and it often leads to inconsistent results.

In this communication we wish to describe a very simple technique which permits the accurate measurement of J-coupling constants from HMQC-J data in a fraction of the time required by other methods. Previously (Wang et al., 1997), we demonstrated how coupling constants could be rapidly extracted from half-height linewidth measurements of unresolved doublets from either TOCSY or NOESY spectra. We have since found that this linewidth measurement protocol can be applied to HMQC-J data (both in the F_1 and F_2 dimensions) with equal accuracy and precision. In particular, when compared to $^3J_{HNH\alpha}$ data predicted from high resolution X-ray structures, the coupling constants measured with this simple linewidth technique have a correlation coefficient of 0.90 and an rmsd error of less than 0.75 Hz.

3.2. Methods

HMQC-J data were obtained for a total of six proteins, all but one of which had a high resolution (< 2.0 Angstroms) X-ray crystal structure. The samples included ^{15}N-labeled (1) *Escherichia coli* thioredoxin, (2) Type III antifreeze protein from Ocean Pout, (3) Type II antifreeze protein from Sea Raven, (4) *Bacillus circulans* xylanase, (5) turkey apo-troponin C (N domain) and (6) calcium-saturated troponin C (E41A-N domain) from turkey. Data were collected at temperatures ranging from 20 - 35 ^{o}C on two different Varian 500 spectrometers, each equipped with a 5 mm triple resonance probe. ^{1}H-^{15}N HMQC-J experiments (Kay and Bax, 1990) were typically performed with a ^{15}N sweepwidth of 2000 Hz and a ^{1}H sweepwidth of 6000 Hz. A total of 2048 complex points were collected along the t_2 domain (^{1}H) and 360 - 400 increments along the t_1 domain (^{15}N). The relaxation delays for these experiments were typically 1.2 - 2.5 s and the refocusing delay was set at between 4.9 - 5.3 ms. Data in both dimensions were zero-filled to create a 4K x 4K data set which was further processed using a shifted sine-bell weighting function (see details below). All ^{1}H-^{15}N HMQC-J spectra were referenced to internal DSS (Wishart et al., 1995).

Each HMQC-J spectrum was assigned on the basis of previously published chemical shift values (with suitable corrections for pH and temperature). In particular, ^{1}H/^{15}N assignments for *E. coli* thioredoxin were based on chemical shifts reported by Wishart (1991) and Chandrasekhar et al. (1991), ^{1}H/^{15}N assignments for the apo and calcium-saturated forms of troponin C (N-domain) were from Gagne et al. (1994), ${}^{1}H/{}^{15}N$ assignments for *B. circulans* xylanase were obtained from those reported by Plesniak et al. (1996), ${}^{1}H/{}^{15}N$ assignments for type III antifreeze protein were from Sonnichsen et al. (1996, personal communication) and the ${}^{1}H/{}^{15}N$ assignments for type II antifreeze protein were supplied by Dr. Wolfram Gronwald (personal communication).

In calculating the coupling constants for each polypeptide, we made use of the following Protein Data Bank entries (see Table 3.1): *E. coli* thioredoxin (2TRX), *B. circulans* xylanase (1BCX) and chicken troponin C (1NCZ). The crystal structure coordinates for type III antifreeze protein were generously provided by Dr. Zongchao Jia of Queens University (personal communication). The coordinate set for chicken troponin C (1NCZ) was chosen over that of turkey troponin C (5TNC) for two reasons: (1) it was of higher quality (better resolution, lower R factor) and (2) the amino acid sequences for the two species are identical, implying that the 3-D structures should also be identical. We also assumed (based on the recent work of Gagne et al., 1997) that the apo- and calcium-loaded forms would have essentially identical structures (save for a small rotation of one bond and a single residue substitution of another) and that one crystal structure would suffice for both the apo and calcium-loaded forms.

${}^{3}J_{HNH\alpha}$ coupling constants for each of the above crystal structures were predicted from the reported backbone ϕ angles using the following equation: ${}^{3}J_{HNH\alpha} = 5.9\cos^{2}\theta - 1.3\cos\theta + 2.2$, where $\theta = |\phi - 60^{o}|$ (see Wang et al., 1997 for more details on the derivation of this equation). Use of other widely used Karplus parameters (Pardi et al., 1984, Vuister and Bax, 1993; Wang and Bax, 1996) led to only minor differences in the overall performance of this method. Note that because a crystal structure is not yet available for the Type II antifreeze protein, we used the ${}^{3}J_{HNH\alpha}$ coupling constants measured from a separate HNHA experiment (Vuister and Bax, 1993) as a proxy for the X-ray crystal values.

3.3. Calculation of ${}^{3}J_{HNH\alpha}$ from $\Delta\nu_{1/2}$

The protocol for determining ${}^{3}J_{HNH\alpha}$ coupling constants for HMQC-J data is very similar to the procedure described earlier for analyzing TOCSY or NOESY spectra (Wang et al., 1997). As stated previously, the spectral resolution prior to zero-filling in both the F_2 and F_1 dimensions should be better than 6.0 Hz/pt. After base-line correction and zero-filling to produce a 4K x 4K data set, a sine-bell weighting function of the form:

$$(3.1) \qquad \sin^2[\pi(t\text{-}sbs)/2sb\}$$

should be applied (where sb and sbs are given in seconds). For Varian Spectrometers running VNMR software (Version 5.1 or higher) sb = sb1 = -0.100 and sbs = sbs1 = -0.066. Note that the negative signs preceding sb and sb1 are used by Varian software to denote a sine-bell "squared" function. For other kinds of spectral processing software, please refer to the conversion formulae provided by Wang et al. (1997).

After processing and assigning the spectrum, one can select traces either from the F_1 (^{15}N) or F_2 (^{1}H) dimension and subsequently determine the half-height linewidth ($\Delta\nu_{1/2}$) for each assigned ^{1}H-^{15}N cross peak. For Varian spectrometers the command "dres" automatically determines $\Delta\nu_{1/2}$ for any given trace. For Bruker spectrometers, $\Delta\nu_{1/2}$ determination takes slightly more effort. By substituting the measured half-width at half-height ($\Delta\nu_{1/2}$) for F_1 (^{15}N) traces into the following equation:

$$(3.2) \qquad {}^3J_{HNH\alpha} = 0.45(\Delta\nu_{1/2}) - MW/20{,}000$$

or the measured half-width at half-height ($\Delta\nu_{1/2}$) for F_2 (^{1}H) traces into this equation:

$$(3.3) \qquad {}^3J_{HNH\alpha} = 0.50(\Delta\nu_{1/2}) - MW/10{,}000$$

the $^3J_{HNH\alpha}$ coupling constant can be determined. Note that $\Delta\nu_{1/2}$ is the half-height linewidth (measured in Hz) of a given ^{1}H-^{15}N resonance and MW is the apparent molecular weight of the protein in daltons. Equations (2) and (3) work well in most situations. However, care must be taken in using the correct molecular weight (i.e. is the polypeptide of interest a monomer or a dimer at NMR concentrations?) and making sure that the temperature of the sample is between 20 °C and 35 °C. Under certain circumstances, the situation can be complicated by the presence of inherently broad linewidths, poor shimming, paramagnetic contaminants or the use of unusually high (>40 °C) or low (< 15 °C) temperatures.

A second approach, which eliminates the problems associated with intrinsic linewidth, temperature or sample differences, can also be used to determine $^3J_{HNH\alpha}$. This method is based on the observation that the narrowest ^1H-^{15}N resonance invariably has a $^3J_{HNH\alpha}$ coupling constant of close to 4.0 Hz. This phenomenon was observed for all six protein samples used in this study and for all 11 polypeptides used in our earlier report (Wang et al., 1997). Using this observation, we have found that $^3J_{HNH\alpha}$ can be determined using either one of the two following equations:

$$(3.4) \quad {}^3J_{HNH\alpha} = 0.45(\Delta\nu_{1/2}) - 0.45(\Delta\nu_{1/2}(\text{min})) + 4.0$$

(for ^{15}N traces)

$$(3.5) \quad {}^3J_{HNH\alpha} = 0.50(\Delta\nu_{1/2}) - 0.50(\Delta\nu_{1/2}(\text{min})) + 4.0$$

(for ^1H traces)

where $\Delta\nu_{1/2}$ is the half-height linewidth (in Hz) of a given ^1H-^{15}N resonance and $\Delta\nu_{1/2}(\text{min})$ is the half-height linewidth (in Hz) of the narrowest ^1H-^{15}N resonance in the spectrum. A small disadvantage to this approach is that coupling constants cannot be determined until after all of the resonance linewidths have been measured and the narrowest line identified. Furthermore, this protocol cannot be applied to the measurement of unstructured peptides or denatured proteins. In these situations the narrowest amide cross peak would likely correspond to a coupling constant of 6 or 7 Hz instead of 4.0 Hz.

3.4. Results

Figure 3.1 illustrates four examples of traces taken through various ^1H-^{15}N cross peaks. As can be seen in these four examples, the $\Delta\nu_{1/2}$ is closely related to the measured $^3J_{HNH\alpha}$ coupling value, with large $\Delta\nu_{1/2}$ values corresponding to large coupling constants and small $\Delta\nu_{1/2}$ values corresponding to small coupling constants. This relationship holds regardless of whether one is measuring in the F_1 (^{15}N) or the F_2 (^1H) dimension. It can be further verified if we plot the relationship between $\Delta\nu_{1/2}$

and the $^3J_{HNH\alpha}$ coupling constant as derived from X-ray data. In Figure 3.2 we illustrate the linear relationship that exists between $\Delta\nu_{1/2}$ (measured along either the ^1H axis or the ^{15}N axis) and $^3J_{HNH\alpha}$ for all measurable resonances from an HMQC-J spectrum of the N domain of turkey apo-troponin C. An excellent fit is obtained for both sets of measurements with correlation coefficients (r) of 0.94 for F_2 (^1H) traces and 0.93 for F_1 (^{15}N) traces. The strong correlation between $\Delta\nu_{1/2}$ and $^3J_{HNH\alpha}$ and the linear relationship observed for these and other examples suggested that a simple equation of the form:

$$(3.6)\quad {}^3J_{HNH\alpha} = m^*\Delta\nu_{1/2} + B$$

(where m is the slope, B is the y intercept and $\Delta\nu_{1/2}$ is the half-height linewidth) could be developed to predict coupling constants from $\Delta\nu_{1/2}$ measurements of HMQC-J spectra. Simulations, using the weighting functions described here and $^1H/^{15}N$ T_2's typical of many mid-sized proteins, confirm that this linear approximation is valid (Figure 3.3).

Obviously, for this method to work effectively it is important to be able to determine the y-intercept (B) independently of the measured linewidths. In Fig. 4 we illustrate how these intercepts can be so determined. In Figure 4.4a the relationship between the "best-fit" y-intercept and the molecular weight of each protein is plotted. Note that the calcium-saturated E41A mutant of troponin C (N-domain) forms a dimer in the presence of calcium (MW_{dimer} = 19.8 kDa), and that type II antifreeze protein (MW_{dimer} = 28.1 kDa) also showed strong evidence of dimer formation under the conditions used in this study. The remaining compounds are known to be monomeric. Also plotted in Figure 4.4b is the relationship between the "best-fit" y-intercept and the half-height linewidth of the narrowest line ($\Delta\nu_{1/2}$(min)). With the exception of *B. circulans* xylanase, which exhibited an unusual linewidth distribution, the straight-line fits to these plots are excellent.

Despite the small problem with xylanase, use of the equations presented here would still allow one to accurately predict the coupling constants of this protein as seen by the data presented in Tables 3.2 and 3.3. These two tables summarize the correlation between these predicted coupling constants (designated as J_{lw} - since they were derived from linewidth measurements) and the coupling constants predicted from the corresponding high resolution X-ray structures (designated as J_{xray}). In assembling these two tables a total of more than 750 coupling constant measurements (387 from ^1H traces; 378 from ^{15}N traces) were made. Both tables clearly show the excellent agreement obtained for both large (28 kDa) and small (7 kDa) proteins using traces from either the ^1H or ^{15}N dimension. Overall, for the six proteins tested, ^1H linewidth measurements yielded an average correlation coefficient of 0.89 and an rmsd from J_{xray} of 0.77 Hz while ^{15}N linewidth measurements yielded an average correlation coefficient of 0.90 and an rmsd from J_{xray} of 0.73 Hz.

3.5. Discussion

In assessing the accuracy of the method presented here, it is important to remember the limitations inherent in comparing X-ray structures with NMR solution structures. Wang and Bax (1996) and Wang et al., (1997) discuss, in some detail, what should be expected in terms of correlation coefficients (r) and RMS deviations between X-ray-derived and experimentally measured coupling constants. Suffice it to say that given the limitations of resolution, thermal motion and signal-to-noise for both techniques, a "perfect" method could probably expect to do no better than $r = 0.97$ and an rmsd = 0.54 Hz between J_{xray} and J_{NMR}. Wang et al. (1997) suggest that a more realistic expectation of ideality would be $r = 0.93$ and an rmsd = 1.04 Hz between J_{xray} and J_{NMR}.

As can be seen from Tables 3.2 and 3.3, our experimentally measured J_{lw} values compare very favorably with the results from our earlier method (Wang et al., 1997) developed explicitly for TOCSY and/or NOESY spectra ($r = 0.89$ and rmsd = 0.85 Hz). They also compare favorably with the original HMQC-J results (based on peak-to-peak measurements) reported by Kay et al. (1989) for Staphylococcal nuclease ($r = 0.89$ and rmsd = 1.01 Hz). The HNHA method of Vuister and Bax (1993) as applied to Staphylococcal nuclease yielded an r value of 0.91 and an rmsd = 0.76 Hz. Later measurements with an expanded data set (Garrett et al., 1994) found that the HNHA experiment produced an agreement between X-ray and NMR results having an r value of 0.78 and an rmsd of 1.42 Hz. The J-modulated COSY approach developed by Billeter et al. (1992) as applied to the 434 repressor protein produced an $r = 0.92$ and an rmsd = 0.76 Hz.

Overall, there is little to distinguish between these methods. Nearly all of the approaches (including the one described here) achieve a level of agreement that is reasonably close to "ideal" (rmsd = 1.04 Hz, r = 0.93). While some methods perform slightly better than ours (0.92 vs. 0.90), it is important to note that our calculations were performed on a substantially larger sample (five to 15 times larger) and a significantly more diverse set of polypeptides (in both size and structure) than any of the other methods. We expect that if the other approaches were applied to a comparably large or diverse data set, their performance would be similarly compromised (see Garrett et al., 1994).

While the above analysis largely confirms that the accuracy and precision of this new approach are as good as any other method currently in use, we believe that the simplicity and rapidity with which J-coupling constants can be determined should make this method particularly appealing to spectroscopists. Specifically, this technique offers four key advantages: (i) the collection and processing of the HMQC-J spectrum only needs to be performed once (as opposed to six or seven times), (ii) the conversion of linewidth measurements to coupling constants can often be done in one's head, (iii) the measurement of linewidths can be accomplished quickly and easily (typically 100 measurements in 30 minutes); and (iv) the $^3J_{HNH\alpha}$ coupling constants can be extracted from both the ^1H and the ^{15}N dimension. This latter point illustrates the robustness of this new approach because it allows one to confirm a coupling constant measurement in two independent ways -- one from a ^{15}N trace and the other from a ^1H trace -- using only a single cross peak. Direct comparisons between the $^3J_{HNH\alpha}$ values derived from the two traces (F_1 and F_2) show that they are highly correlated (r = 0.95) and this further suggests that if a trace in one dimension is obscured or distorted, then a trace in the other dimension (if it is not distorted or obscured) could be used to extract a coupling constant with a high degree of confidence.

While there are many positive aspects to this simple approach to $^3J_{HNH\alpha}$ coupling constant determination, there are at least a few limitations that merit further discussion. One obvious shortcoming is the fact that, in order for this method to work, the HMQC-J spectra must be collected and processed in a very specific manner.

While this can be a hindrance, the reprocessing of a previously collected HMQC-J spectrum (with modern computers) should only take a few seconds. On the other hand, competing methods based primarily on computer-aided curve fitting (Kay et al., 1989; Billeter et al., 1992; Goodgame and Geer, 1993; Vuister and Bax, 1993) are much more flexible and do not typically constrain the user to follow special collection and processing conditions. Another limitation of this linewidth-based technique arises from the fact that it can be sensitive to conditions that affect linewidths, but which may not necessarily affect $^3J_{HNH\alpha}$ coupling constants. Such variables as temperature, paramagnetic contaminants, solvent viscosity, non-uniform segmental motion, dimerization events, intermediate exchange events, spectral overlap and decoupler distortion can all affect linewidth measurements -- yet these phenomenon often have little to do with a protein's average backbone structure or its $^3J_{HNH\alpha}$ coupling constants. Consequently, these common sources of lineshape perturbation or distortion can potentially lead to incorrect $^3J_{HNH\alpha}$ values. A third limitation lies with the potential difficulties associated with determining the "y-intercept". As seen with the *B. circulans* xylanase example, it is sometimes possible to introduce a systematic error (up to 0.5 Hz) in coupling constant measurements through an incorrect determination of the "y-intercept" or correction factor. Care, therefore, must be taken to ensure that this correction factor is consistent with what is known about the molecule and that it yields a range of $^3J_{HNH\alpha}$ values typical for proteins (between 3 Hz and 10 Hz). Despite these possible limitations, we have found that this technique has worked very well for every protein so far tested.

To summarize, we have described a novel method that allows $^3J_{HNH\alpha}$ coupling constants to be rapidly determined from simple linewidth measurements in either the ^1H or ^{15}N dimension of HMQC-J cross peaks. This new method makes use of the linear relationship between $^3J_{HNH\alpha}$ and half-height linewidths ($\Delta\nu_{1/2}$) of appropriately processed NMR spectra. We believe this approach offers several advantages over other previously described heteronuclear techniques for extracting $^3J_{HNH\alpha}$ coupling constants. In particular, it is simple, quick, accurate (having an rmsd of less than 0.8 Hz), easy to learn, applicable to both small and large proteins, independent of any requirement for specialized hardware, and independent of the spectrometer make, size or type. We believe that if this simple concept of linewidth measurement is widely adopted, it could make quantitative coupling constant measurements far simpler and far easier to use in analyzing the solution conformation of peptides, proteins, polynucleotides and other biomolecules via NMR.

References

Bax, A., Vuister, G. W., Grzesiek, S., Delaglio, F., Wang, A. C., Tschudin, R. and Zhu, G. (1994) *Meth. Enzymol.* 239, 79-105.

Billeter, M., Neri, D., Otting, G., Qian, Y. Q. and Wuthrich, K. (1992) *J. Bomol. NMR* 2, 257-274.

Billeter, N., Neri, D., Otting, G., Qian, Y. Q. and Wuthrich, K. (1992) *J. Biomol. NMR* 2, 257-274.

Chandrasekhar, K., Krause, G., Holmgren, A. and Dyson, H. J. (1991) *FEBS Lett.* 284, 178-183.

Forman-Kay, J. D., Gronenborn, A. M., Kay, L. E., Wingfield, P. T. and Clore, M. G. (1990) *Biochemistry* 29, 1566-1572/

Gagne, S. M., Li, M. X. and Sykes, B. D. (1997) *Biochemistry* 36, 4386-4392.

Gagne, S. M., Tsuda, S., Li, M. X., Chandra, M., Smillie, L. B. and Sykes, B. D. (1994) *Protein Science* 3, 1961-1974.

Garrett, D. S., Kuszewski, J., Hancock, T. J., Lodi, P. T., Vuister, G. W., Gronenborn, A. M. and Clore, G. M. (1994) *J. Magn. Reson. B.* 104, 99-103.

Goodgame, M. M. and Geer, S. M. (1993) *J. Magn. Reson. A.* 102, 246-248.

Katti, S. K., LeMaster, D. M. and Eklund, H. (1990) *J. Mol. Biol.* 212, 167-184.

Kay, L. E. and Bax, A. (1990) *J. Magn. Reson.* 86, 110-126.

Kay, L. E., Brooks, B., Sparks, S. W., Torchia, D. A. and Bax, A. (1989) *J. Am. Chem. Soc.* 111, 5488-5490.

Pardi, A., Billeter, M. and Wuthrich, K. (1984) *J. Mol. Biol.* 180, 741-751.

Plesniak, L. A., Wakarchuk, W. W. and Mcintosh, L. P. (1996) *Portein Science* 5, 1118-1135.

Satyshur, K. A., Pyzalska, D., Rao, S. T., Greaser, M. and Sundaralingam, M. (1994) *Acta Cryst. D* 50, 40-49.

Sonnichsen, F. D., DeLuca, C. I., Davies, P. L. and Sykes, B. D. (1996) *Structure* 4, 1325-1337.

Sonnichsen, F. D., Sykes, B. D., Chao, H. and Davies, P. L. (1993) *Science* 259, 1154-1157.

Vuister, G. W. and Bax, A. (1993) *J. Am. Chem. Soc.* 113, 7772-7777.

Wakarchuk, W. W., Cambell, R. L., Sung, W. L., Davoodi, J. and Yaguchi, M. (1994) *Protein Science* 3, 467-474.

Wang, A. C. and Bax, A. (1996) *J. Am. Chem. Soc.* 118, 2483-2494.

Wang, Y., Nip. A. M. and Wishart, D. S. (1997) *J. Biomol. NMR (in press)*

Weisemann, R., Ruterjans, J., Schalbe, H., Schleucher, J., Bermel, W. and Griesinger, C. (1994) *J. Biomol. NMR* 4, 231-240.

Wishart, D. S. (1991) Ph.D. Thesis, Yale University, New Haven.

Wishart, D. S., Bigam, C. G., Yao, J., Abildgaard, F., Dyson, H. J., Oldfield, E., Markley, J. L. and Sykes, B. D. (1995) *J. Biomol. NMR* 6, 135-140.

Table 3.1. Listing of high resolution X-ray structures used in calculating J_{xray} values.

Protein	Accession	Resolution	R-factor	Reference
Thioredoxin (*E. coli*)	2TRX	1.68 Å	0.165	Katti, S.K. et al. (1991)
Troponin C (Apo)	1NCZ	1.80 Å	0.19	Satyshur, K.A. et al. (1994)
Troponin C (E41A)	1NCZ	1.80 Å	0.19	Satyshur, K.A. et al. (1994)
Xylanase (*B. circulans*)	1BCX	1.80 Å	0.163	Wakarchuk, W.W. et al. (1994)
Antifreeze protein (III)	N/A	1.25 Å	0.14	Z. Jia (personal communication)

Table 3.2. Summary of results obtained using linewidth analysis for determination of $^3J_{HNH\alpha}$ coupling constants (^{1}H dimension only).

Protein	MW (daltons)	No. of points	r (J_{xray} vs J_{lw})	rmsd (J_{xray} vs J_{lw})
Antifreeze protein (III)	6860	50	0.92	0.71
Troponin C (Apo)	9900	69	0.94	0.59
Thioredoxin (E. coli)	11880	56	0.91	0.82
Troponin C (E41A)	19800	60	0.88	0.73
Xylanase (B. circulans)	20400	108	0.84	0.83
Antifreeze protein (II)	28000	44	0.91	0.95

Table 3.3. Summary of results obtained using linewidth analysis for determination of $^3J_{HNH\alpha}$ coupling constants (^{15}N dimension only).

Protein	MW (daltons)	No. of points	r (J_{xray} vs J_{lw})	rmsd (J_{xray} vs J_{lw})
Antifreeze protein (III)	6860	52	0.93	0.67
Troponin C (Apo)	9900	67	0.93	0.58
Thioredoxin (E. coli)	11880	50	0.93	0.74
Troponin C (E41A)	19800	57	0.83	0.82
Xylanase (B. circulans)	20400	108	0.88	0.71
Antifreeze protein (II)	28000	44	0.91	0.92

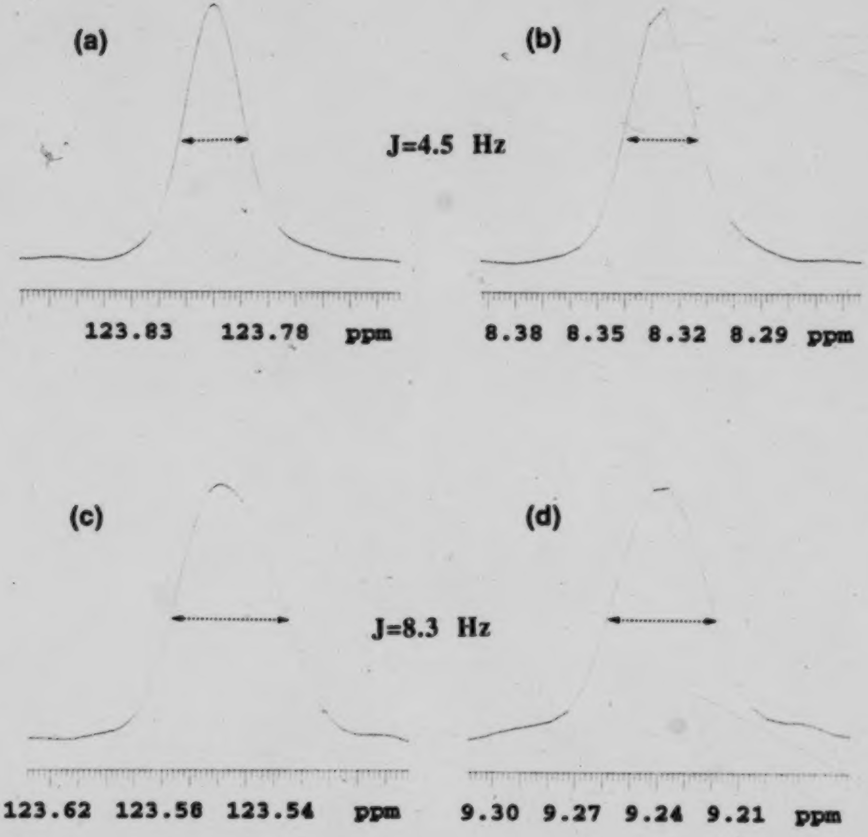


Figure 3.1. Four examples of traces taken through ^{1}H-^{15}N cross peaks from an HMQC-J spectra of *B. circulans* xylanase. Illustrated in (a) and (b) are the traces of Lys^{154} taken through the ^{1}H (F_2) dimension and ^{15}N (F_1) dimension respectively. The measured half-height linewidth in (a) was 12.1 Hz, while the measured half-height linewidth in (b) was 13.4 Hz. Illustrated in (c) and (d) are the traces of Val^{98} taken through the ^{1}H (F_2) dimension and ^{15}N (F_1) dimension respectively. The measured half-height linewidth in (c) was 19.7 Hz, while the measured half-height linewidth in (d) was 18.8 Hz. The $^3J_{HNH\alpha}$ value (in Hz) as predicted from high resolution X-ray data is indicated in each figure. Note that broad peaks are associated with large coupling constants while narrow peaks are associated with small coupling constants.

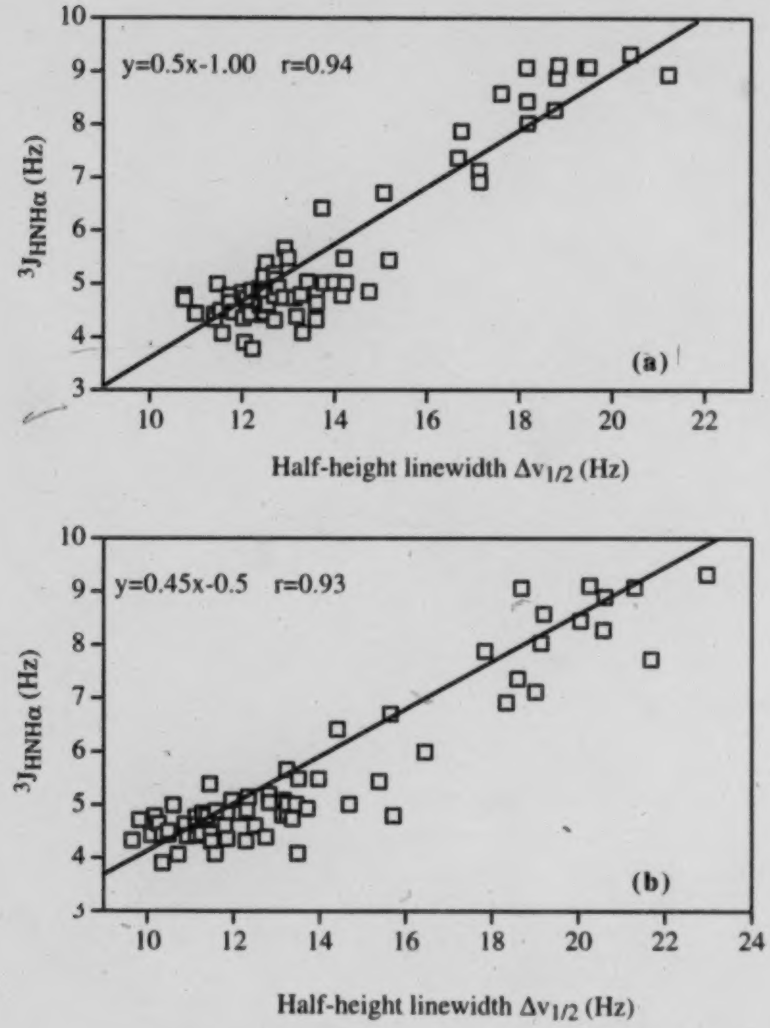


Figure 3.2. Relationship between $\Delta\nu_{1/2}$ and the predicted $^3J_{HNH\alpha}$ (from X-ray data) for apo-troponin C as determined from (a) ^{1}H (F_2) traces and (b) ^{15}N (F_1) traces respectively. The equations for the "best-fit" line derived from the molecular weight based approach and the correlation coefficients (r) are shown in the top left corner of each graph. Note that the superimposed curve is a "best fit" line for all of the data (390 points) and all of the proteins (six) and so, for any given protein, there may be slight systematic deviations at certain extrema.

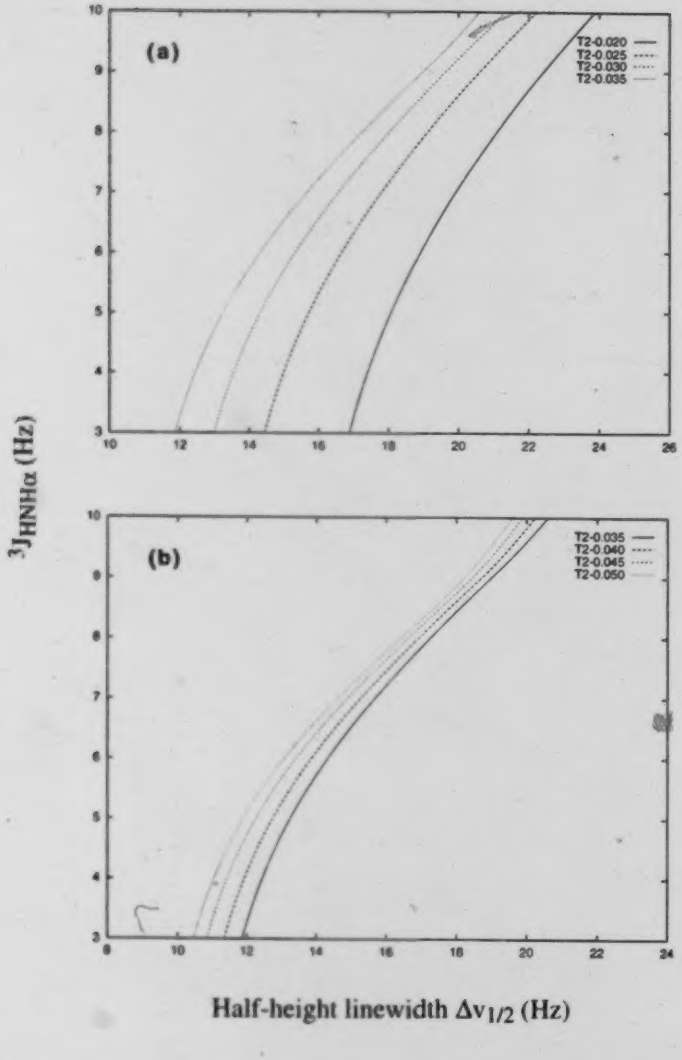


Figure 3.3. Computer simulation of the relationship between $\Delta\nu_{1/2}$ and $^3J_{HNH\alpha}$ coupling constants for (a) ^{1}H traces and (b) ^{15}N traces using T_2 values typical of mid-sized proteins and the sine-bell processing parameters suggested in the manuscript.

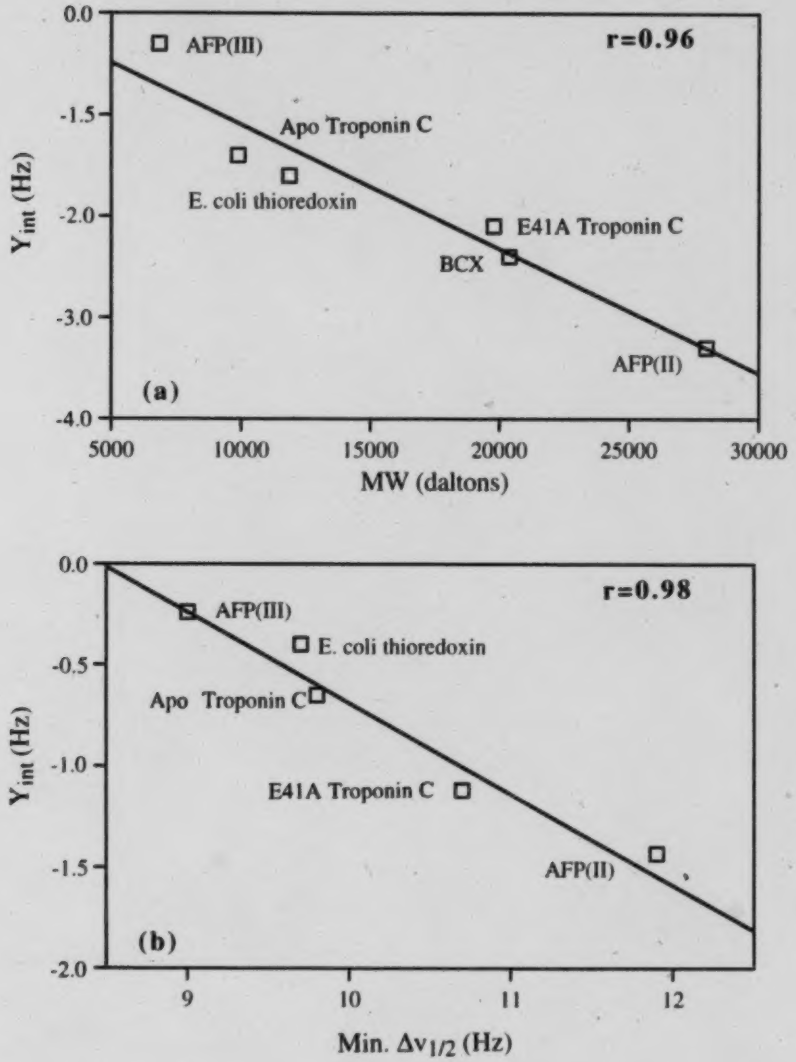


Figure 3.4. (a) Relationship between the "best-fit" y-intercept and the molecular weight of each of the test proteins derived from ^{1}H linewidth measurements and (b) the relationship between the "best-fit" y-intercept and the linewidth of the narrowest amide resonance for each test protein derived from ^{15}N linewidth measurements. The correlation coefficient (excluding BCX in the lower Figure) is given in the top right corner of each graph.

Chapter 4*

NMR Solution Structures of Oxidized and Reduced Bacteriophage T4 Glutaredoxin

4.1. Introduction

Glutaredoxins are small, ubiquitous, structurally conserved family of proteins typically containing fewer than one hundred amino acids. They have been identified and sequenced in essentially all classes of organisms, including bacteriophage (Gvakharia et al., 1996a), bacteria (Aslund et al., 1994; Gvakharia et al, 1996b; Vlamis-Gardikas, 1997), yeast (Gan et al., 1990; Luikenhuis et al., 1998), plants (Morell et al., 1995; Sha et al., 1997; Szederkenyi et al., 1997), invertebrates (Ludemann et al., 1998; Ebel et al., 1997) and mammals (Papayannopoulos et al., 1989; Hopper et al., 1989; Padilla et al., 1996; Park and Levine, 1996). All glutaredoxins characterized to date contain a common dithiol/disulfide redox active-site motif composed of four critical residues: C-[PV]-[FYW]-C. The primary biological role of glutaredoxin is to function as a hydrogen donor for ribonucleotide reductase, a key enzyme in DNA synthesis (Holmgren, 1989). Other biological functions include the catalysis of dehydroascorbate to ascorbate (Wells et al., 1990; Martensson & Meister, 1991), the reactivation of DNA-binding activity of certain nuclear factors (Bandyopadhyay et al., 1998), the acceleration of *in vitro* protein folding rates (Lundstrom-Ljung and Holmgren, 1995), and the regulation and/or maintenance of HIV-1 protease activity (Davis et al., 1997). Glutaredoxins share many structural and functional similarities to another important member of the thioltransferase superfamily -- the thioredoxins (Trx). However, thioredoxins and glutaredoxins carry out their respective redox activities using fundamentally different mechanisms. In particular, the thioredoxin system uses NADPH, thioredoxin reductase, and thioredoxin; while the glutaredoxin system uses NADPH, glutathione reductase, glutathione, and glutaredoxin. Thioredoxin is reduced by thioredoxin reductase while glutaredoxin is reduced by glutathione (Holmgren, 1989).

Bacteriophage T4 Grx is a rather unique member of the thioltransferase superfamily. While sequence comparisons suggest that it belongs to the glutaredoxin family (Nikkola et al., 1991), T4 Grx's level of sequence similarity to any other

---

*This chapter will be submitted by Wang, Y. & Wishart, D.S. to *Eur. J. Biochem.*

glutaredoxin rarely exceeds 20%, making it something of an evolutionary outlier. Furthermore, T4 glutaredoxin is the only glutaredoxin with a valine substitution in the highly conserved second position of the C-P-[FYW]-C active site motif. These global and local sequence differences likely contribute to its unusual hybrid functionality as well. In particular, T4 Grx appears to function both as a glutaredoxin and a thioredoxin. Recent experiments have shown that it can be reduced not only by glutathione but also by thioredoxin reductase (Nikkola et al, 1993). Indeed, it was because of this functional hybridism that T4 glutaredoxin was originally known as bacteriophage T4 thioredoxin (Berglund et al., 1970; Sjoberg and Holmgren, 1972; Soderberg et al., 1978; Eklund et al., 1984; Joelson et al., 1990). It is only recently (Nikkola et al., 1991; Eklund et al., 1992) that this protein has become formally known as T4 glutaredoxin.

The small size (87 residues), good stability and "missing link" evolutionary character of T4 glutaredoxin has made it the subject of a number of detailed structural and functional studies. To date, four X-ray structural studies of T4 Grx have been published, including the original report of the oxidized form at 2.8 Å resolution (Soderberg et al., 1978), the refined 2.0 Å resolution structure (Eklund et al., 1992); a crystallographically stable T4 Grx mutant at 1.45 Å resolution (Eklund et al., 1992) and, most recently, an active site mutant (Val15-Gly; Tyr16-Pro) of the reduced form at 2.4 Å resolution (Ingelman et al., 1995). These crystallographic studies have revealed a protein with high level of ordered secondary structure (> 60%), a good mix of helices and beta sheets and only one structurally non-essential disulfide bond. These favorable structural qualities have also made T4 Grx the subject of several protein folding studies using stopped-flow fluorescence and CD spectroscopy (Borden and Richards, 1990a, 1990b).

Despite these extensive X-ray crystallographic and UV spectroscopic studies on the structure and folding of T4 Grx, no solution structure has yet been reported for this molecule. Furthermore, the X-ray structure of the reduced wild-type form has, so far, proven to be elusive. Here, we wish to report the complete ^{1}H assignments and high resolution solution structures of both the oxidized and reduced forms of wild-type T4 Grx as determined by ^{1}H NMR spectroscopy. We also report the results of a detailed structural comparison between these two molecular forms and discuss the consequences that certain observed structural, electrostatic and dynamic differences may have in the function of this unique, "hybrid" protein. It is hoped that this work will eventually allow more detailed studies to be undertaken of the folding pathway(s), stability and, ultimately, the mechanism of action for T4 Grx and its relatives.

4.2. Materials and Methods

Sample preparation. T4 Grx was prepared from the *E. coli* strain DL39, transformed with a temperature inducible pUC18 plasmid containing both the *E. coli* Trx (thioredoxin) gene and the phage T4 Grx gene (nrdC) tandemly ligated into an XbaI site (LeMaster, 1986; LeMaster and Richards, 1988). Cells were grown at 30 °C in LB medium until an OD_{600} of 1.0 was reached, at which point the temperature was shifted to 42 °C to induce both *E. coli* Trx and T4 Grx production. After 3 hours of growth, the cells were pelleted by centrifugation (3000g, 10 min) and resuspended in a 50-mM sodium phosphate, pH 7.0 buffer and lysed by sonication (5 min). The protein-containing lysate was separated from the cellular debris by centrifugation (12000g, 20 min), whereupon the lysate was then subjected to ammonium sulfate precipitation (4 °C, overnight). The fraction precipitating at 35-75% was isolated by centrifugation and resuspended in a 50-mM sodium phosphate, pH 7.0 buffer. After elution through a G-75 super sephadex size exclusion column, the final product was separated and purified by a DE52 anion exchange column using a 50-400 mM NH_4HCO_3 gradient. The T4 Grx typically elutes at 125 mM while *E. coli* thioredoxin normally elutes in at 250 mM NH_4HCO_3. While modest yields (15 mg/L) of T4 Grx were obtainable when cells were grown in rich media, attempts to prepare ^{15}N and ^{13}C labelled T4 Grx in minimal medium were not successful. For NMR studies, a lyophilized T4 Grx sample was dissolved in a pH 6.0 buffer containing 20 mM potassium phosphate and 150 mM NaCl in 99.99% D_2O or 90%H_2O/10%D_2O yielding a final protein concentration of 2-3 mM. A small quantity (0.1 mM) of 2,2-dimethyl-2-silapentane-5-sulfonic acid (DSS) was added to all samples as a chemical shift reference (Wishart et al., 1995). The reduced form of T4 Grx was prepared by adding excess (10 mM) dithiothreitol (DTT) to the NMR sample after which it was blanketed with argon for 30 minutes and then sealed with an airtight rubber septum.

NMR spectroscopy. All ^{1}H NMR experiments were performed on Varian Unity 600 MHz and VXR 500 MHz NMR spectrometers equipped with 5-mm triple resonance probes. TOCSY experiments were carried out using the basic pulse sequence proposed by Bax and Davis (Bax & Davis, 1985). Acquisition times were generally set to 0.2 seconds, relaxation delays were 2.0 seconds and spin-lock (MLEV-17) mixing times ranged from 50 to 100 milliseconds. NOESY (Jeener et al., 1979; Kumar et al., 1980) spectra were collected essentially identically to the TOCSY spectra, with mixing times ranging from 150 to 300 milliseconds. Phase-sensitive DQF-COSY spectra were collected using standard protocols (Shaka & Freeman, 1983; Rance et al., 1984). All two-dimensional spectra were collected using 256 t_1 increments and spectral widths of 6000 Hz in both dimensions. To distinguish between overlapping signals, additional TOCSY and NOESY spectra were collected at different temperatures (15 ° C - 35 °C). For all NMR experiments, the sample temperature was maintained to within ± 0.1°C. Water suppression was achieved by solvent presaturation during the relaxation delay period. Prior to Fourier transformation, each data matrix was zero-filled to 4K x 4K complex points and multiplied by an approximate 90°-shifted sine-bell squared weighting function in both dimensions. Complete ^{1}H NMR chemical shift assignments were carried out using well-established procedures (Wuthrich, 1986; Clore and Gronenborn, 1987) in which individual spin systems were initially identified from TOCSY and DQF-COSY spectra and sequential assignments were obtained from NOESY spectra. To identify slowly exchanging amide protons, a lyophilized T4 Grx sample was dissolved in D_2O at 10 °C in 20-mM potassium phosphate, pH 6.0. The sample was purged with argon for 2 min and then inserted into a preshimmed spectrometer whereupon a complete NOESY spectrum was acquired over 5 hours. χ_1 torsion angles and stereospecific methylene proton assignments were extracted from DQF-COSY and NOESY spectra (Smith et al., 1991). $^3J_{HNH\alpha}$ coupling constants were obtained using line-width measurements from both TOCSY and NOESY spectra (Wang et al., 1997).

Structure Generation. Structure calculations were carried out using the same protocols for both reduced and oxidized proteins. Interproton distance restraints were derived from the assigned NOE peaks of a NOESY spectrum acquired with a mixing time of 150 milliseconds. The assigned NOE intensities (measured by volume integration) were classified into four groups: strong, medium, weak and very weak, corresponding to interproton distance restraints of 1.8-2.8 Å, 1.8-4.0 Å, 1.8-5.0 Å and 1.8-6.0 Å respectively. Upper distance limits for distances involving methyl protons and nonstereospecificlly assigned methylene protons were corrected for center averaging (Wuthrich et al., 1983), by adding 0.5 Å to the upper distance limits for NOEs involving methyl protons. $^{3}J_{HNN\alpha}$ coupling constants were measured from both TOCSY and NOESY spectra using the line-width measurement techniques of Wang et al. (1997). A total of 46 ϕ backbone torsion angles restraints were derived from 68 $^{3}J_{HNH\alpha}$ coupling constants measured from both TOCSY and NOESY spectra for both oxidation states of protein. Specifically; for those residues in well-defined β-sheet regions ($^{3}J_{HNN\alpha}$ > 8.5 Hz), ϕ angle restraints were set to -120 ± 30°; for those residues in well defined helical regions, ϕ angles were estimated from an appropriately parameterized Karplus equation (Wang and Bax, 1996) and assigned an uncertainty of ± 15°. The remaining 22 coupling constants, for which the ϕ angle could not be explicitly determined, were used directly as coupling constant restraints and assigned an uncertainty of ± 1.5 Hz. Backbone ψ dihedral angle restraints were obtained from an analysis of $d_{N\alpha/d\alpha N}$ ratios. For $d_{N\alpha/d\alpha N}$ ratios less than one, the ψ angle restraint was set to 120 ± 100°. For $d_{N\alpha}/d_{\alpha N}$ ratios greater than one, the ψ angle restraint was set to -30 ± 110° (Gagne et al, 1994). Throughout our calculations, all ω angles were set to 180 ± 10° except for cis proline 66 which was set to 0 ± 10°. Backbone hydrogen bonds were identified from previously obtained secondary structure information together with data from amide exchange measurements. χ_1 torsion angles restraints from DQF-COSY and NOESY spectrum were set to 180°, -60° or 60° with an uncertainty of ± 30°. Each hydrogen bond was defined using two distance restraints, $d_{O\text{-}H}$=1.8-2.4 Å and $d_{O\text{-}N}$=2.7-3.5 Å. Three distance restraints, $d_{S(i)\text{-}S(j)}$=2.02 ± 0.05 Å, $d_{S(i)\text{-}C\beta(j)}$=2.99 ± 0.05 Å and $d_{C\beta(i)\text{-}S(j)}$=2.99 ± 0.05 Å were used to define the disulfide bond between residues 14 and 17 for oxidized T4 Grx. Force constants for NOE-derived distance restraints were set to 50 $\text{kcalmol}^{-1}\text{Å}^{-2}$, while dihedral angle force constants were initially set at 5 $\text{kcalmol}^{-1}\text{rad}^{-2}$ during the high-temperature dynamics run and increased to 200 $\text{kcalmol}^{-1}\text{rad}^{-2}$ during the annealing stage.

In total 1050 interproton distance restraints (401 intra-residue, 198 sequential, 133 medium range, 257 long range, and 66 hydrogen bond restraints), 46 ϕ angle restraints, 56 ψ angle restraints, 86 ω angle restraints, 13 χ_1 angle restraints, 22 $^3J_{HNH\alpha}$ coupling constant restraints and 208 ^{1}H chemical shift restraints were used in the calculation for the oxidized protein. The numbers of restraints used for the reduced form were approximately the same as for the oxidized form (given the high overall spectral similarity). However, careful comparison between the NOESY spectra of reduced and oxidized T4 Grx yield an additional 15 NOE restraints for the reduced form. For the first stage of structure generation, only NOE derived interproton distance restraints were included as experimental input. Initially, fifty structures (for each redox form) were generated using the simulated annealing protocol of Nilges et al. (1988) as implemented in the X-PLOR package (version 3.8.5 - Brunger, 1992). Twelve thousand high-temperature steps (60 ps at 1000 K) followed by 6000 cooling steps (30 ps, final temperature of 100K) were used in this protocol. After this step, several ambiguous long-ange NOE assignments were then clarified by analyzing the initial NMR structures and the known X-ray structures. Further refinement was carried out using a similar protocol but with 6000 high-temperature steps (30 ps) and 4000 cooling steps (20 ps). Those structures which exhibited no inter-proton distance restraint violations greater than 0.5 Å were accepted and used as input structures for the second stage of refinement, in which torsion angle restraints were introduced. For the final stage of structural refinement, those structures which exhibited no inter-proton distance or torsion angle restraint violation greater than 0.5 Å or 5°, respectively, were accepted and further refined against $^3J_{HNH\alpha}$ coupling constant restraints (Garrett et al., 1994) and proton chemical shift restraints (Kuszewski et al., 1995). During this final refinement stage, 800 steps of conjugate-gradient minimization were performed. The force constants for $^3J_{HNN\alpha}$ coupling constant restraints and proton chemical shift restraints were 1.0 kcalmol^{-1}Hz^{-2} and 7.5 kcalmol^{-1}ppm^{-2}. Throughout the refinement process, the force constants for interproton distance restraints and dihedral angle restraints were set to 50 kcalmol^{-1}Å^{-2}, 200 kcalmol^{-1}rad^{-2} respectively. The quality of the final ensemble of structures was assessed with PROCHECK-NMR (version 3.4.4, Laskowski et al., 1996). MOLMOL (Koradi et al., 1996) was used to visualize all protein structures and to calculate RMSD values.

Measurement of Thiol Ionization by Ultraviolet Absorbance. T4 Grx thiol ionization constants were measured using the protocol described by Dyson et al.

(1997). Specifically 1 mL of a 2 mM T4 Grx sample was reduced using 10X molar excess DTT in a 0.1 mM EDTA, 50 mM phosphate buffer (pH 6.5) under an argon atmosphere. After reduction, excess DTT was removed by dialysis (3 X 500 mL) against the same buffer (previously saturated with argon for 30 min). The dialysis apparatus was maintained under an argon atmosphere throughout the 15 hour dialysis period. After dialysis the T4 Grx sample was diluted to a concentration of ~30 μM in a 0.1 mM EDTA, pH 6.0, 100 mM potassium phosphate buffer. Thiol ionization was monitored by measuring the protein absorbance at 240 nm on a Pharmacia Biotech, Ultraspec 3000 UV/Vis spectrophotometer. The sample pH was adjusted up or down by adding small aliquots of 2 M NaOH or 2 M HCl as needed. The concentration of T4 Grx was determined from the absorbance at 280 nm using the extinction coefficients (1.0 A_{280} unit corresponds to 1.52 mg/ml) proposed by Joelson et al. (1990).

4.3. Results and Discussion

Sequential assignment and secondary structure. ^{1}H chemical shift assignments of oxidized and reduced T4 Grx were completed according to well-established protocols (Wuthrich et al., 1982; Wuthrich, 1986). Spin systems were identified using both TOCSY spectra (acquired with different mixing times) and DQF-COSY spectra. Sequential assignments were then obtained by analyzing NOESY spectra for short through-space NOE connectivities of NH, C$^\alpha$H and C$^\beta$H protons. Ambiguities arising from chemical shift degeneracy were resolved by recording TOCSY or NOESY spectra at different temperatures. Complete ^{1}H chemical shift assignments for oxidized and reduced T4 glutaredoxin are listed in Tables 4.1 and 4.2 respectively.

The secondary structure of T4 Grx was determined from an analysis of C$^\alpha$H proton chemical shifts (Wishart et al., 1992; Wishart and Sykes, 1994), backbone $^3J_{HNH\alpha}$ coupling constants, amide proton exchange rates, and a qualitative analysis of the sequential, medium-range and long-range backbone NOEs (Wuthrich, 1986). Both the reduced and oxidized forms of T4 Grx contains three α–helices extending from residues 14 to 27, 44 to 55, and 78 to 86, as indicated by the characteristic $d_{NN}(i,i+1)$, $d_{NN}(i,i+2)$, $d_{\alpha N}(i,i+3)$, and $d_{\beta N}(i,i+3)$ NOE connectivities, small (<6 Hz)

$^3J_{HNH\alpha}$ coupling constants, slowly exchanging amide protons, and upfield $C^{\alpha}H$ proton chemical shifts (Figure 4.1). A portion of a NOESY spectrum (for oxidized T4 Grx) showing the $d_{NN}(i,i+1)$ connectivity in these helices is presented in Figure 4.2. T4 Grx also contains a four-stranded β-sheet as shown in Figure 4.3. The first β-strand (β1) extends from residues 2 to 7, the second (β2) from 30 to 36, the third (β3) from 64 to 70 and the fourth (β4) from 73 to 77. These β-strands are characterized by stretches of strong sequential $d_{\alpha N}(i,i+1)$ NOE connectivities, long-range $d_{NN}(i,j)$, $d_{\alpha N}(i,j)$, $d_{\alpha\alpha}(i,j)$ backbone NOE connectivities, large (> 8.5 Hz) $^3J_{HNH\alpha}$ coupling constants, slowly exchanging amide protons, and downfield α-proton chemical shifts (Figure 4.1).

NMR solution structures of T4 Grx. As described previously, an initial set of fifty structures was generated for both oxidized and reduced T4 Grx. These structures were further refined against dihedral angle, $^3J_{HNH\alpha}$ coupling constant and proton chemical shift restraints. The acceptance criteria for the final set of T4 Grx structures (both for the oxidized and reduced forms) was as follows: no interproton distance restraint violation greater than 0.5 Å, no dihedral angle restraint violation grater than 5°, no rms deviation from ideal bond length grater than 0.02 Å, no rms deviation from ideal bond angle greater than 2.0°, no $^3J_{HNH\alpha}$ coupling constant restraint violation greater than 1.5 Hz, and no more than 25 proton chemical shift restraint violations greater than 0.5 ppm (i.e. $|\delta_{obs}-\delta_{calc}|>0.5$ ppm) for any given structure. For some structures, several rounds of refinement were required to meet the above criteria. A total of 30 out of the initial set of 50 structures satisfied the above criteria and were accepted into the final ensemble of T4 Grx solution structures.

Ribbon diagrams of the final restrained minimized mean structures of oxidized and reduced T4 Grx are shown in Figure 4.4. As can be seen from this figure, T4 Grx is composed a four-strand mixed parallel and anti-parallel β-sheet surrounded by three α-helices. The structure also contains two loop regions from residues 9 to 14 and 56 to 63. The redox-active site (Cys14-Val15-Tyr16-Cys17) is located on the N-terminal side of the first helix. Like other thioredoxins and glutaredoxins (Xia et al., 1992), this redox-active site is solvent accessible on one side of the molecule, but solvent inaccessible on the other. Close inspection of the active site structure shows that the side-chains of Tyr7, Tyr16 and Met65 effectively block any approaches to Cys17. On the other hand, Cys14 remains partly exposed in both oxidized and reduced T4 Grx. As shown in Figure 4.4, the side-chains of Tyr7, His12, Tyr16, and Met65 actually form a cleft around the active site thiols. Mutational analysis show that these residues play an important role in the biological activity of T4 Grx (Nikkola et al., 1991; Joelson et al., 1990). As with all other thioredoxins and glutaredoxins, T4 Grx has a cis proline located near its active site (Eklund et al., 1991; Nikkola et al., 1991). The cis configuration of the Met65-Pro66 peptide bond, which is present in both oxidized and reduced T4 Grx, was easily identified from the characteristic $d_{\alpha\alpha}$(i, i+1) NOE connectivities seen in our NOESY spectra (Wuthrich et al., 1984). The unusual cis conformation of Pro66 has been proposed to facilitate glutathione binding into the active site (Nikkola et al., 1991). It is also believed to play a key role in stabilizing the tertiary structure of T4 Grx during the latter stages of folding (Borden and Richards, 1990b).

The quality and precision of each structural ensemble was assessed with the program PROCHECK-NMR (Laskowski et al., 1996). For both the oxidized and reduced set of structures, the overall G-factor was 0.11 ± 0.03 and 0.12 ± 0.03 respectively. With regard to the ϕ/ψ distribution of backbone dihedral angles (Ramachandran et al., 1963; Laskowski et al., 1996), more than 83% of residues for both sets of 30 structures were found to lie within the most allowed region. The average hydrogen bond energies were calculated to be 0.78 ± 0.08 and 0.81 ± 0.10 kcal/mol, and the number of bad contacts per 100 residues was determined to be 1.9 ± 1.2 and 2.1 ± 1.4 for the oxidized and reduced forms respectively. These parameters indicate that the quality of our T4 Grx NMR structures are comparable to a 2.0 Å X-ray structure. Stereo diagrams of the 30 superimposed backbone traces for both oxidized and reduced T4 Grx are shown in Figure 4.5. As can be seen from this figure, the global fold of T4 Grx is very well-defined with relatively few structural deviations for either of the two sets of structures. More specifically, the backbone RMSD is just 0.71 Å for the oxidized form and 0.70 Å for the reduced form. These RMSD values fall to just 0.59 Å for the oxidized form and 0.60 Å for the reduced form when residues in the flexible loops (9-14, 56-63, and 71-73) are excluded from the calculation. To further assess the status of these three flexible loops, local backbone RMSD values for overlapping sets of tripeptides as well as for the full length structures were calculated and plotted (Figure 4.6) for both oxidized and reduced T4 Grx. This figure shows that these three loops are locally disordered in both reduced and oxidized T4 Grx. A complete summary of the NMR structural statistics for T4 glutaredoxin is provided in Table 4.3.

Structural comparison between oxidized and reduced T4 Grx. A best-fit backbone superposition of the minimized mean NMR structures of oxidized and reduced T4 Grx is presented in Figure 4.7. Careful comparison between these two structures as well as additional comparisons between the observed NOE connectivities, $^3J_{HNH\alpha}$ coupling constants, chemical shifts and location of slowly exchanging amide protons reveals a very high degree of similarity between the two redox forms. This observation is consistent with an earlier X-ray crystallographic study of an active site T4 Grx mutant (Val15-Gly and Tyr16-Pro) that showed only minor structural perturbations occurring after reduction of the disulfide bond (Ingelman et al., 1995). This is also consistent with NMR data previously reported for the oxidized and reduced forms of three related thiol-transferases. In particular, detailed NMR studies of human thioredoxin (Forman-Kay et al., 1991; Qin et al., 1994), human glutaredoxin (Sun et al., 1998) and *E. coli* thioredoxin (Dyson et al., 1990; Jeng et al., 1994) all reveal that the overall folds for both the reduced and oxidized forms are essentially identical.

Although a high level of structural similarity exists for both oxidized and reduced T4 Grx, there are at least three notable differences that merit further discussion. The first and most obvious difference concerns the geometry of the active site cysteines. Upon reduction, the Sγ atoms of Cys14 and Cys17 move apart quite substantially. In particular, the average distance between the two Sγ atoms for the ensemble of 30 reduced T4 Grx structures was measured to be 4.05 ± 0.87 Å. In contrast, the distance between the two Sγ atoms in the oxidized form was just 2.02 ± 0.05 Å. This degree of separation is consistent with data obtained for other members of the glutaredoxin/thioredoxin family. In particular, the distance between the Cys23 Sγ and Cys26 Sγ atoms in reduced human Grx was found to be 4.1 Å (Sun et al., 1998) while the distance between the Cys32 Sγ and Cys35 Sγ atoms in reduced *E. coli* Trx was measured at 3.8 Å (Jeng et al., 1994). It is particularly interesting that earlier X-ray studies (Ingelman et al., 1995) of a reduced T4 Grx active-site mutant found that the Sγ atoms of Cys14 and Cys17 were separated by just 2.6 Å -- some 1.4 Å less than what we have measured for wild-type T4 Grx and some 1.2 to 1.5 Å less than what has been measured for other wild-type thioredoxins/glutaredoxins. Interestingly, the short Sγ separation distance reported by Ingelman et al. is consistent with results previously reported for an NMR structure of an active-site mutant of reduced human Trx, where the Sγ separation was found to be a relatively short 3.1 Å (Qin et al., 1994). These latter results may indicate that the introduction of a mutation into the active site could limit the flexibility of this region, particularly on reduction of the macrocyclic Cys-Val-His-Cys ring. This change in entropy (relative to the wild-type reduced state) could be partially responsible for the limited activity observed for these active site mutants. Conversely, this result also suggests that a significant conformational change in the cysteine Sγ positions (or accommodation of these changes) is critical to the proper functioning of the wild-type glutaredoxins/thioredoxins.

The second significant difference between reduced and oxidized T4 Grx concerns their chemical shifts (αH and HN) and coupling constants ($^3J_{HNH\alpha}$). As might be expected, reduction of T4 Grx leads to moderate (>0.1 ppm) α-proton chemical shift changes for a number of residues proximal to the redox active site. These include Gly6, Tyr7, Val15, Tyr16, Cys17, Asp18, Asn19, Asn35, Ile36, Thr64, Gln67, His75 and Gly78. Even larger chemical shift changes (>0.3 ppm) are observed for amide protons belonging to Tyr7, Cys14, Val15, Ile36, Gln60, Met65 and His75. A plot of the αH and HN chemical shift differences between oxidized and reduced T4 Grx is shown in Figure 4.8. It is well known that chemical shifts offer valuable structural information (Wishart et al., 1991 and 1994; Williamson et al, 1992 and 1995; Osapay and Case, 1991 and 1994) but their structural consequences are often difficult to discern. In a very general sense, α-protons are primarily affected by ring currents and backbone dihedral angles, while amide protons are most affected by ring currents and hydrogen bond interactions (Pardi et al., 1983; Wishart et al., 1991; Blanchard et al., 1997). Inspection of either of the high resolution T4 Grx X-ray structures or our high resolution NMR structures allows one to identify which residues (and atoms) could possibly be affected by ring currents. Visual comparisons, distance measurements and chemical shift calculations (TOTAL) indicate that the α-protons of both Tyr7 and Asn35 are influenced by Phe33, while the α-protons of Asn19 and His75 are affected by Phe79 and Phe69, respectively. This suggests that the rather substantive (>0.3 ppm) changes seen in the active site Cys17 and Asp18 α-proton shifts (as well as those of Thr64 and Ile36) upon reduction are likely due to cumulative backbone dihedral angle changes of >20° (Wishart et al., 1991; Wishart and Nip, 1998). On the other hand, substantive upfield amide shift changes seen in Cys14 and Val15 after reduction can probably be attributed to the loss (or weakening by ~1 kcal/mol) of two hydrogen bonds (Wishart et al., 1991). We suggest this is likely the case because neither of these residues is close enough to be perturbed by a ring (as measured in our structures) and neither has undergone a substantive backbone change (as judged by the lack of α-proton shift changes for these residues). While reduction in T4 Grx likely leads to the loss of two hydrogen bonds in the active site, hydrogen bond interactions between Tyr7 and Ile36 seem to be strengthened to a similar (or compensatory) degree as judged by their large downfield amide chemical shift changes. Similar chemical shift changes (both in magnitude and direction) for residues in the vicinity of the active site have also been reported for *E. coli* thioredoxin (Dyson et al, 1988 and 1989; Chandrasekhar et al., 1994). These chemical shift changes are largely mirrored in the changes seen for $^3J_{HNH\alpha}$ coupling constants between reduced and oxidized T4 Grx (Table 4.5). In particular, significant (>0.8 Hz) changes are seen in the $^3J_{HNH\alpha}$ values for Val4, Tyr7, Gln29, Asn35, Ile36 and Ile61 after reduction. The $^3J_{HNH\alpha}$ coupling constant changes measured for Tyr7 and Ile36 suggest that these two residues likely undergo a backbone conformational change to increase their amide hydrogen bonding potential.

The third and final difference of note concerns the differential line broadening observed between the oxidized and reduced T4 Grx. A listing of the measured half-height linewidths for most amide TOCSY cross peaks in both oxidized and reduced T4 Grx is given in Table 4.4. On average, the half-height linewidth of amide resonances in oxidized T4 Grx are 1.2 Hz larger than those of the reduced form. Previous work by our group has shown that it is possible to use NMR linewidth measurements to accurately determine the oligermization status (or apparent molecular weight) of peptides and proteins at concentrations significantly higher (>1 mM) than what could be measured via ultra centrifugation (Wang et al., 1997; Wishart and Wang, 1998). Specifically, we have shown that a simple linear relationship exists between a macromolecule's average half height linewidth and its apparent molecular weight. Using the data from Table 4.4 and the equations given in Wang et al. (1997), we calculate that the apparent molecular weight of reduced T4 Grx is 11,500 daltons whereas the apparent molecular weight of oxidized T4 Grx is 19,500 daltons. This suggests that reduced T4 Grx (at the concentrations used in this study) exists as a monomer while oxidized T4 Grx likely exists as a dimer. This interesting result is actually consistent with previously published work on two related glutaredoxins/thioredoxins. Specifically, sedimentation equilibrium measurements and detailed NMR backbone dynamics studies of *E. coli* Grx-1 (Kelley et al, 1997) and *E. coli* Trx (Stone et al., 1993) have shown that these two proteins tend to exist as dimers in the oxidized form and monomers in the reduced form. X-ray studies of human thioredoxin show that this molecule exists as a dimer in both oxidized and reduced states (Weichsel et al., 1996) while Gronenborn et al. (1999) have shown that this molecule exists as a monomer in solution. Whether this dimer to monomer transition is of biological or functional relevance is difficult to ascertain. Certainly cellular concentrations of Grx or Trx are far below the concentrations used in these NMR experiments and so it would appear unlikely that dimerization is necessary for function. Meanwhile, recent controversies surrounding presumptive dimer-monomer transitions in interleukins (Rajarathnam et al., 1994 and 1997) and other small proteins suggest that this is an issue worthy of further investigation.

Comparison with T4 Grx X-ray structures. Four different X-ray structures for T4 Grx or T4 Grx mutants have been reported over the past twenty years (Soderberg et al., 1978; Eklund et al., 1992; Ingelman et al., 1995). Reported crystallographic resolutions range from 2.8 Å for the 1978 structure to 1.45 Å (R-factor = 0.175) for the 1992 structure. For this study we have chosen to compare our NMR solution structures of wild-type T4 Grx to the 1.45 Å X-ray structure of an active-site T4 Grx mutant reported by Eklund et al. (1992). A best-fit backbone atom superposition of this 1.45 Å X-ray structure with the two NMR solution structures (reduced and oxidized) is shown in Figure 4.8. The backbone atomic RMSD between each restraint minimized NMR structure and the X-ray structure is 1.51 Å for the oxidized form, and 1.47 Å for the reduced form (Table 4.3). While the overall folds of both NMR structures and the X-ray structure are very similar, some differences exist between them.

Among the most obvious are the differences in backbone dihedral angles. A comparison between the observed $^3J_{HNH\alpha}$ coupling constants with the predicted values (Wang and Bax, 1996) derived from the 1.45 Å X-ray structure is given in Table 4.5. For at least four residues (His12, Val68, Phe69 and His75) differences between predicted and measured $^3J_{HNH\alpha}$ coupling constants exceed 2.0 Hz. In the X-ray structure, ϕ angles for both Val68 and Phe69 were measured to be -120°. In both NMR structures, we find they are actually -85° and -155° respectively. Similarly, the ϕ angles for His12 and His75 were measured to be -143° and -77° in the X-ray structure, but were found to be -175° and -60 ° in both NMR structures. Given the inherent limitations of coupling constant measurement and structure determination by NMR (Wang and Bax, 1996; Wang et al., 1997), it is possible that these differences simply reflect experimental inaccuracies or structural averaging effects. However, it is also possible that these discrepancies indicate very real differences between the solution and solid state or, equally likely, differences between the mutant and wild-type structures. Given the generally good agreement between the observed and predicted $^3J_{HNH\alpha}$ coupling constants for all other residues in T4 Grx, we are inclined to believe that these differences are a consequence of real structural perturbations.

In addition to these scattered main-chain dihedral angle differences, there is a notable difference in the 56-61 loop region between the NMR and the X-ray structures. In this region the solution structure is not particularly well-defined due to a low density of restraints. The X-ray structure clearly indicates that the side-chains of Leu51 and Leu52 in helix $\alpha 2$ should be quite close (<5 Å) to the side-chains of Arg57, Thr59 and Asn60 in this particular loop. However, no such interaction could be found in NOESY spectra collected with a variety of different mixing times. The lack of NOEs does not necessarily suggest that these side-chain interactions are absent in solution, but it does suggest that the loop is either highly flexible or that it has a dynamically averaged conformation that is different than the X-ray structure. Interestingly, a recent study of human glutaredoxin has shown that the same (structurally equivalent) loop exhibits a similar displacement relative to that seen in the X-ray structure of the highly homologous pig liver glutaredoxin (Sun et al., 1998). Additional comparisons of the X-ray and NMR structures of *E. coli* thioredoxin and human thioredoxin also reveal notable structural displacements in the same loop regions.

While main chain differences between the NMR and X-ray structures do exist, it is also important to emphasize that there is still very good agreement throughout most regions of T4 Grx, including the active site side chains. Comparison of the NMR structure's active site (oxidized) shows that the side chain orientations of Tyr7, His12, Val15 and Met65 are well defined and in a good agreement with the 2.0 Å structure of Eklund et al. (1992). Specifically, the χ_1 torsion angles for Tyr7, His12 and Met65 are -64°, -176°, -62° and -68° in NMR structure and -61°, -171°, -54° and -75° in the X-ray structure. The χ_1 angle of Tyr16 is also well-defined in the NMR structure but differs slightly from the X-ray structure (-103° vs. -81°). The side chain orientations of Lys13, Cys14 and Cys17 are not sufficiently well-defined in the NMR structure to make meaningful comparisons.

Comparison with other glutaredoxins and thioredoxins. More than a dozen three-dimensional structures of thioredoxins and glutaredoxins have been reported over the past 25 years. These studies are summarized in Table 4.6. From this table it is evident that the atomic structures of six evolutionary distinct members of the thioredoxin/glutaredoxin superfamily (*E. coli* Trx, *E. coli* Grx, human Trx, human Grx, calf thymus Grx, and bacteriophage T4 Grx) have been solved by X-ray crystallography and/or NMR spectroscopy. Despite significant sequence differences, every study published to date has shown that these proteins share the same overall fold, with either an N-terminal $\beta1/\alpha1/\beta2/\alpha2/\beta3$ or a $\beta2/\alpha2/\beta3$ structural unit joined to a C-terminal $\beta4/\beta5/\alpha4$ unit by a loop involving the third helix ($\alpha3$). For all six structures, the $\beta2$, $\beta3$, $\beta4$, and $\beta5$ strands form a central, continuous, twisted pleated sheet with helices $\alpha1$ and/or $\alpha2$ located on one side of the sheet and helices $\alpha3$ and $\alpha4$ located on other side. This central β-sheet is highly conserved. Indeed, a backbone superposition of this four-stranded β-sheet for all six highly divergent proteins (minimum pairwise sequence identity < 15%) yields a RMSD value of only 0.85 Å. This level of conservation is quite remarkable and it suggests that the central β-sheet scaffold must play a key role in the folding and stability of all members of this ubiquitous protein family.

In addition to the high level of structural conservation seen in the central β-sheet, there is also an equally high level of structural conservation in their active site loops. A backbone superimposition of the Cys-Xaa-Xaa-Cys redox active site for all six solved glutaredoxin/thioredoxin molecules yields an RMSD value of just 0.143 Å. This high level of structural similarity in both thioredoxin and glutaredoxin active sites indicates just how precise the backbone geometry must be in order to confer the redox activity seen in this class of proteins. It is interesting that this structural similarity persists even though the T4 Grx active site sequence (Cys-Val-Tyr-Cys) is quite different than the usual thioredoxin (Cys-Gly-Pro-Cys) or glutaredoxin (Cys-Pro-Tyr[Phe]-Cys) active site sequences. While the active site loop is structurally well conserved, it is important to emphasize that the character and disposition of key side chains around this active site is not. The nature and consequences of these differences will be discussed in more detail below.

Measurement of T4 Grx thiol ionization by UV spectroscopy. Among various members of the glutaredoxin and thioredoxin family, the stability of the disulfide bond is known to vary widely (Takahashi and Creighton, 1996). This stability has been directly correlated with the pKa value (which tends to be below 7.0) of the most accessible of the two cysteine thiols. On the other hand, the pKa value for the more buried cysteine tends to be very high (above 9.0) and very similar for both glutaredoxins and thioredoxins (Gan et al., 1990; Meiyal et al., 1991; Jeng and Dyson. 1996; Dyson et al., 1997; Sun et al., 1998) Normally the thiol pKa values of the more exposed cysteines in glutaredoxins are lower (pKa <5) than in thioredoxins (pKa >6.5). Because T4 Grx displays characteristics of both a thioredoxin and a glutaredoxin, we decided to investigate the disulfide stability and thiol ionization constants of T4 Grx more thoroughly. However, the weak ^{1}H NMR signal and high degree of spectra overlap observed near the two active site cysteines in T4 Grx precluded pKa measurement by NMR spectroscopy - consequently we choose to use an alternative experimental technique, UV spectroscopy.

Indeed, for quite some time it has been known that the difference in the UV extinction coefficient between a thiol and its thiolate anion is about 4000 $M^{-1}cm^{-1}$ at 240 nm (Polgar, 1974). Recently, this easily measurable difference has been used to determine the pKa values of thiols in several thiol-transferases including DsbA

(Nelson and Creighton, 1994), *E. coli* thioredoxin (Takahashi and Creighton, 1996), and an *E. coli* thioredoxin mutant (Dyson et al., 1997). Here we report the results of similar thiol titration experiments for T4 Grx. As can be seen in Figure 4.9, two titration events are evident, one with a midpoint of 6.8 and a $\Delta\varepsilon 240$ of 4000 $M^{-1}cm^{-1}$ between pH 6.4 - 7.6 and a second, less obvious one, with a midpoint around pH 9.5. The first titration event likely corresponds to the ionization of Cys14, which is the more accessible thiol. The last titration event most certainly corresponds to the ionization of Cys17, which is more buried. Because T4 Grx is now classified as a glutaredoxin, a somewhat lower pKa (< 5.0) for Cys14 has actually been predicted for quite some time (Gan and Wells, 1987; Sandberg et al., 1990; Eklund et al., 1992; Ingelman et al., 1995). However, our titration results indicate that this is not the case. Indeed they suggests that T4 Grx has an active site titration chemistry that is actually more akin to a thioredoxin than a glutaredoxin.

Traditionally, the greater solvent accessible surface area for active site thiols has been used to explain the lowered pKa values in glutaredoxins relative to thioredoxins (Gan and Wells, 1987; Sun et al., 1998). To test this hypothesis, the pKa and the atomic accessible surface area of the exposed cysteines for T4 Grx, *E. coli* Trx (as a representative thioredoxin), and pig liver Grx (as a representative glutaredoxin) were analyzed using VADAR (Wishart et al., 1995). The results are listed in Table 4.7. Despite having clearly differing pKa values, the atomic accessible surface area for the most exposed cysteine in each protein is almost indistinguishable. This suggests that the exposed surface area may actually have little to do with the differing glutaredoxin/thioredoxin pKa values. On the other hand, a good correlation between the calculated electrostatic potential and observed thiol pKa values has recently been demonstrated by Sun et al (1998).

In particular, the unusually low pKa values for the nucleophilic thiols of Grx and Trx appear to arise from the presence of a large, positive local electrostatic potential. The exposed cysteines for T4 Grx, *E. coli* Trx and pig Grx as well as their surrounding positively charged amino acids are shown in Figure 4.10. It can been seen from this figure that there are relatively few basic amino acids surrounding the *E. coli* Trx and T4 Grx thiols, but substantially more basic (positively charged) residues surround the active site thiols in pig liver Grx. The large, local positive electrostatic potential found in pig liver Grx (and the low pKa for its nucleophilic Cys) likely arises from the proximity to five positively charged residues: Lys19, Agr26, Lys27, Arg67, and Arg71, all of which are highly conserved in the glutaredoxin family. On the other hand, T4 Grx only has two positively charged residues (His12 and Lys13) close to its redox-active site. Since histidine can only be positively charged at low pH values, it is unlikely to affect the ionization of a thiol at neutral pH. Furthermore, the side chain of Lys13 extends on the opposite direction with that of the Cys14, so it is unlikely to affect the ionization potential of the Sγ of Cys14 in any significant way. Structural analyses also show that the terminal amino groups on the His12 and Lys13 side chain are well separated from Sγ of Cys14. So it is extramely unlikely that these residues could form a "salt bridge" to Cys14 and perturbe it's pKa value. Consequently the paucity of positive charge at the T4 Grx active site appears to be the only feature that could account for the unexpectedly high pKa of Cys14.

4.4. Conclusion

The importance of glutaredoxins and thioredoxins in facilitating vital enzymatic processes, combined with their small size and extensive sequence diversity (Holmgren, 1989; Holmgren and Bjornstedt, 1995; Holmgren and Aslund 1995; Dyson, 1995) makes studies of their structure, function, dynamics and folding pathways attractive. T4 Grx represents a very unique member of this superfamily and we believe that by completing the ^{1}H NMR assignments and solution structure determination of both the reduced and oxidized forms of T4 Grx we may have opened the way to more detailed studies directed at understanding the folding pathway(s), stability and, ultimately, the mechanism of action for this unique protein.

References

Aslund, F., Ehn, B., Miranda-Vizuete A., Pueyo, C. and Holmgren, A. (1994) *Proc. Natl. Acad. Sci. U.S.A.* 91, 9813-9817.

Bandyopadhyay, S., Starke, D. W., Mieyal, J. J. and Gronostajski, R. M. (1998) *J. Biol. Chem.* 273, 392-397.

Bax, A. and Davis, D. G. (1985) *J. Magn. Reson.* 65, 355-360.

Berglund, O. and Sjoberg, B. M. (1970) *J. Biol. Chem.* 245, 6030-6035.

Blanchard, L., Hunter, C. N. and Williamson, M. P. (1997) *J. Biolmol. NMR* 9, 389-395.

Borden, K. L. B. and Richards, F. M. (1990a) *Biochemistry* 29, 8207-8210.

Borden, K. L. B. and Richards, F. M. (1990b) *Biochemistry* 29, 3071-3077.

Brunger, A.T. (1992) X-PLOR Version 3.1. A System for X-ray Crystallography and NMR, Yale University Press, New Haven, CT

Bushweller, J. H., Billeter, M., Holmgren, A. and Wuthrich, K. (1994) *J. Mol. Biol.* 235, 1585-1597.

Chandrasekhar, K., Krause, G., Holmgren, A. and Dyson H. J. (1991) *FEBS Lett.* 284, 178-183.

Chandrasekhar, K., Patricia, Campbell, P. A., Jeng, M. F., Holmgren, A. and Dyson, H. J. (1994) *J. Biomol. NMR* 4, 411-432.

Clore, G. M. and Gronenborn, A. M. (1987) *Protein Eng.* 1, 275-288.

Davis, D. A., Newcomb, F. M., Starke, D. W., Ott, D. E., Mieyal, J. J. and Yarchoan, R. (1997) *J. Biol. Chem.* 272, 25935-25940.

Dyson, H. J. (1995) *Methods Enzymol.* 252, 293-306.

Dyson, H. J., Holmgren, A. and Wright, P. E. (1988) *FEBS Lett.* 228, 254-258.

Dyson, H. J., Holmgren, A. and Wright, P. E. (1989) *Biochemistry* 28, 7074-7087.

Dyson, H. J., Feng, M.F., Tennant, L. L., Slay, I., Lindell, M., Cui, D. S., Kuprin, S. and Holmgren, A. (1997) *Biochemistry* 36, 2622-2236.

Dyson, J. J., Gippert, G. P., Case, D. A., Holmgren, A. and Wright, P. E. (1990) *Biochemistry* 29, 4129-4136.

Ebel, T., Middleton, J. F. S., Frisch, A. and Lipp, J. (1997) *J. Biol. Chem.* 272, 3042-3048.

Eklund, H., Cambillau, C., Sjoberg, B-M., Holmgren, A., Jornvall, H., Hoog, J-O. and Branden, C-I. (1984) *EMBO J.* 3, 1443-1449.

Eklund, H., Gleason, F. K. and Holmgren, A. (1991) *Proteins* 11, 13-28.

Eklund, H., Ingelman, M., Soderberg, B-O., Uhlin, T., Nordlund, P., Nikkola, M., Joelson, T. and Petratos, K. (1992) *J. Mol. Biol.* 228, 596-618.

Forman-Kay, J. D., Clore, G. M., Wingfield, P. T. and Gronenborn, A. M. (1991) *Biochemistry* 30, 2685-2698.

Gagne, S. M., Tsuda, S., Li, M. X., Chandra, M., Smillie, L. B. and Skyes, B. D. (1994) *Prot. Sci.* 3, 1961-1974.

Gan, Z. R. and Wells, W. W. (1987) *J. Biol. Chem.* 262, 6704-6707.

Gan, Z. R., Polokoff , M. A., Jacobs, J. W. and Sardana, M. K. (1990) *Biochem. Biophys. Res. Commum.* 168, 944-951.

Garrett, D. S., Kuszewski, J., Hancock, T. J., Lodi, P. J. Vuister, G. W., Gronenborn, A. M. and Clore, G. M. (1994) *J. Magn. Reson.*, Series B 104, 99-103.

Gronenborn, A. M., Clore, G. M. and Louis, J. M. (1999) *Prot. Sci.* 8, 426-429.

Gvakharia, B. O., Hanson, E., Koonin, E. K. and Mathews, C. K. (1996) *J. Biol. Chem.* 271, 15307-15310.

Gvakharia, B. O., Koonin, E. K. and Mathews, C. K. (1996) *Virology* 226, 408-411.

Holmgren, A. (1985) *Annu. Rev. Biochem.* 54, 237-271.

Holmgren, A. (1989) *J. Biol. Chem.* 264, 13963-13966.

Holmgren, A. and Aslund, F. (1995) *Methods Enzymol.* 252, 283-291.

Holmgren, A. and Bjornstedt, M. (1995) *Methods Enzymol.* 252, 199-208.

Holmgren, A., Soderberg, B-O., Eklund, H. and Branden, C. I. (1975) *Proc. Natl. Acad. Sci. U.S.A.* 72, 2305-2309.

Hopper, S., Johnson, R. S., Vath, J. E. and Biemann, K. (1989) *J. Biol. Chem.* 264, 20438-20447.

Ingelman, M., Nordlund, P. and Eklund, H. (1995) *FEBS Lett.* 370, 209-211.

Jeener, J., Meier, B. H., Bachmann, P. and Ernst, R. R. (1979) *J. Chem. Phys.* 71, 4546-4553.

Jeng, M. F. and Dyson, H. J. (1996) *Biochemistry* 34, 611-619.

Jeng, M. F., Campbell, A. P., Begley, T., Holmgren, A., Case, D. A., Wright, P. E. and Dyson, H. J. (1994) *Structure* 2, 853-868.

Joelson, T., Sjoberg, B-M. and Eklund, H. (1990) *J. Biol. Chem.* 265, 3183-3188.

Katti, S. K., LeMaster, D. M. and Eklund, H. (1990) *J. Mol. Biol.* 212, 167-184.

Katti, S. K., Robbins, S. H., Yang, Y. and Wells, W. W. (1995) *Prot. Sci.* 4, 1998-2005.

Kelley, J. J., Caputo, T. M., Eaton, S. F., Laue, T. M. and Bushweller, J. H. (1997) *Biochemistry* 36, 5029-5044.

Koradi, R., Billeter , M. and Wuthrich, K. (1996) *J. Mol. Graph.* 14, 29-32.

Kumar, A., Ernst, R. R. and Wuthrich, K. (1980) *Biochem. Biophys. Res. Commun.* 95, 1-6.

Kuszewski, J., Gronenborn, A. M. and Clore, G. M. (1995) *J. Magn. Reson.*, Series B 107, 293-297.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. and Thornton, J. M. (1993) *J. Appl. Crystallogr.* 26, 283-291.

LeMaster, D. M. (1986) *J. Virol.* 59, 759-760.

LeMaster, D. M. and Richards, F. M. (1988) *Biochemistry* 27, 142-150.

Ludemann, H., Dormeyer, M., Sticherling, C., Stallmann, D., Follmann, H. and Krauth-Siegel, R. L. (1998) *FEBS Lett.* 431, 381-385.

Luikenhuis, S., Perrone, G., Dawes, I. W. and Grant, C. M. (1998) *Molecular Biology of the Cell* 9, 1081-1091.

Lundstrom-Ljun, J. and Holmgren, A. (1995) *J. Biol. Chem.* 270, 7822-7828.

Martensson, J. and Meister, A. (1991) *Proc. Natl. Acad. Sci. U.S.A.* 88, 4656-4660.

Mieyal, J. J., Starke, D. W., Gravina, S. A. and Hocevar, B. A. (1991) *Biochemistry* 30, 8883-8891.

Morell, S., Follmann, H. and Haberlein, I. (1995) *FEBS Lett.* 369, 149-152.

Nelson, J. W. and Creighton, T. E. (1994) *Biochemistry* 33, 5974-5983.

Nikkola, M., Gleason, F., Saarinen, M., Joelson, T., Bjornberg, O. and Eklund, H. (1991) *J. Biol. Chem.* 266, 16105-16112.

Nikkola, M., Gleason, F. K. and Eklund, H. (1993) *J. Biol. Chem.* 268, 3845-3849.

Nilges, M., Clore, G. M. and Gronenborn, A. M. (1988) *FEBS Lett.* 229, 317-324.

Osapay, K. A. and Case, D. A. (1991) *J. Am. Chem. Soc.* 113, 9436-9444.

Osapay, K. A. and Case, D. A. (1994) *J. Biomol. NMR* 4, 215-230.

Padilla, C. A., Bajalica, S., Lagercrantz, J. and Holmgren, A. (1996) *Genomics* 32, 455-457.

Papayannopoulos, I. A., Gan, Z. R., Wells, W. W. and Biemann K. (1989) *Biochem. Biophys. Res. Commum.* 159, 1448-1454.

Park, J. B. and Levine, M. (1996) *Biochemical Journal* 315, 931-938.

Polgar, L. (1974) *FEBS Lett.* 38, 189-190.

Qin, J., Clore, G. M. and Gronenborn, A. M. (1994) *Structure* 2, 503-522.

Rajarathnam, K., Kay, C. M., Clark, L. I. and Sykes, B. D. (1997) *Methods in Enzymol.* 287, 89-105.

Rajarathnam, K., Sykes, B. D., Kay, C. M., Dewald, B., Geiser, T. Baggiolini, M. and Clark, L. I. (1994) *Science* 264, 90-92.

Ramachandran, G. N., Ramakrishnan, C. and Sasisekharan, V. (1963) *J. Mol. Biol.* 7, 95-99.

Rance, M., Sorensen, O. W., Bodenhausen, G., Wagner, G., Ernst, R. R. and Wuthrich, K. (1984) *Biochem. Biophys. Res. Commum.* 117, 479-485.

Sha, S., Minakuchi, K., Higaki, N., Sato, K., Ohtsuki, K., Kurata, A., Yoshikawa, H. Kotaru, M., Masumura, T., Ichihara, K. and Tanaka, K. (1997) *J. Biochem.* 121, 842-848.

Shaka, A. J. and Freeman, R. (1983) *J. Magn. Reson.* 51, 161-169.

Sjoberg, B. M. and Holmgren, A. (1972) *J. Biol. Chem.* 247, 8063-8068.

Sodano, P., Chary, K. V. R., Bjornberg, O., Holmgren, A., Kren, B., Fuchs, J. A. and Wuthrich, K. (1991) *Eur. J. Biochem.* 200, 369-377.

Sodano, P., Xia, T. H., Bushweller, J. H., Bjornber, O., Holmgren, A., Billeter, M. and Wuthrich, K. (1991) *J. Mol. Biol.* 221, 1311-1324.

Soderberg, B. O., Sjoberg, B. M., Sonnerstam, U. and Branden, C. I. (1978) *Proc. Natl. Acad. Sci. U.S.A.* 75, 5827-5830.

Soderberg, B. O., Sjoberg, B.M., Sonnerstam, U. and Brundon, C. I. (1978) *Proc. Nat. Acad. Sci., U.S.A.* 75, 5827-5830.

Stone, M. J., Chandrasekhar, K., Holmgren, A., Wright, P. E. and Dyson, H. J. (1993) *Biochemistry* 32, 426-435.

Sun, C. H., Berardi, M. J. and Bushweller, J. H. (1998) *J. Mol. Biol.* 280, 687-701.

Szederkenyi, J., Komor, E. and Schobert, C. (1997) *Planta* 202, 349-56.

Takahashi, N. and Creighton, T. E. (1996) *Biochemistry* 35, 8342-8353.

Vlamis-Gardikas, A., Aslund, F., Spyrou, G., Bergman, T. and Holmgren, A. (1997) *J. Biol. Chem.* 272, 11236-11243.

Wang, C. A. and Bax, A. (1996) *J. Am. Chem. Soc.* 118, 2483-2494.

Wang, Y., Nip, A. M. and Wishart, D. S. (1997) *J. Biomol. NMR* 10, 373-382.

Weichsel, A., Gasdaska, J. R., Powis, G. and Montfort, W. R. (1996) *Structure* 15, 735-751.

Wells, W. W., Pen, X. D., Yang, Y. and Rocque, P. A. (1990) *J. Biol. Chem.* 265, 15361-15364.

Williamson, M. P., Asakura, T., Nakamura, E. and Demura, M. (1992) *J. Biomol. NMR* 2, 83-98.

Williamson, M. P., Kikuchi, J. and Asakura, T. (1995) *J. Mol. Biol.* 247, 541-546.

Wishart, D. S. and Nip, A. M. (1998) *Biochem. Cell Biol.* 76, 153-163.

Wishart, D. S. and Sykes, B. D. (1994) *Methods Enzymol.* 239, 363-392.

Wishart, D. S. and Sykes, B. D. and Richards, F. M. (1992) *Biochemistry* 31, 1647-1651.

Wishart, D. S. and Wang, Y. (1998) *J. Biomol. NMR* 11, 329-336.

Wishart, D. S., Kondejewki, L. H., Semchuk, P. D., Kay, C. M., Hodges, R. S. and Sykes, B. D. (1995) *Techniques in Protein Chemistry. VI*, 451-457.

Wishart, D. S., Sykes, B. D. and Richards, F. M. (1991) *J. Mol. Biol.* 222, 311-333.

Wishart, D. S., Bigam, C. G., Yao, J., Abildgaard, F., Dyson, H. J., Oldfield, E., Markley, J. L. and Sykes, B. D. (1995) *J. Biomol. NMR* 6, 135-140.

Wishart, D. S., Willard, L. and Sykes, B. D. (1995) VADAR 1.1 - University of Alberta (ftp://redpoll.pharmacy.ualberta.ca)

Wuthrich, K. (1986) *NMR of Proteins and Nucleic Acids.* Wiley Inter-Science, New York.

Wuthrich, K., Billeter, M. and Braun, W. (1983) *J. Mol. Biol.* 169, 949-961.

Wuthrich, K., Billeter, M. and Braun, W. (1984) *J. Mol. Biol.* 180, 715-740.

Wuthrich, K., Wider, G., Wagner, G. and Braun, W. (1982) *J. Mol. Biol.* 155, 311-319.

Xia, T. H., Bushweller, J. H., Sodano, P., Billeter, M., Bjornber, O., Holmgren, A. and Wuthrich, K. (1992) *Prot. Sci.* 1, 310-321.

Table 4.1. ^{1}H chemical shift* assignments for oxidized bacteriophage T4 Grx

residues	NH	αH	βH	others
M1		4.39	2.21, 2.35	γCH_2 2.64, 2.55; εCH_3 2.01
F2	8.76	4.66	3.34, 2.87	δH 7.43; εH 7.25; ξH 7.48
K3	8.73	5.25	1.94, 1.58	γCH_2 1.53, 1.51; δH_2 1.99, 1.73; εH_2 2.90, 2.90
V4	9.03	4.87	2.33	γCH_3 0.91, 0.86
Y5	9.61	5.19	*2.64, 3.19*	δH 7.11; εH 6.57
G6	8.47	4.10, 3.87		
Y7	7.27	5.38	*2.86, 3.62*	δH 7.31; εH 6.83
D8	8.73	4.68	2.91, 2.84	
S9	7.40	4.14	4.61, 4.41	
N10	8.38	4.50	3.22, 3.03	
I11	8.10	4.07	2.11	γCH_2 1.20, 0.93; γCH_3 1.00; δCH_3 0.83
H12	8.88	4.51	*2.83, 2.75*	$\delta^2 H$ 5.90; $\delta^1 H$ 8.54
K13	8.09	4.24	1.44, 1.68	γCH_2 1.74, 1.74; δH_2 1.99, 1.99; εH_2 3.15, 3.15
C14	9.18	4.77	3.39, 3.39	
V15	9.13	3.92	1.85	γCH_3 0.97, 0.76
Y16	9.19	4.21	3.27, 3.27	δH 7.28; εH 7.06
C17	8.49	4.40	2.88, 2.78	
D18	7.95	4.31	*2.74, 2.90*	
N19	8.82	4.26	1.85, 1.85	δNH_2 6.98, 5.83
A20	7.93	4.18	1.51	
K21	7.45	3.26	1.54, 1.54	γCH_2 1.73, 1.22; δH_2 1.94, 1.94
R22	8.13	4.03	1.92, 1.92	γCH_2 1.68, 1.75; δH_2 3.24, 3.24; εH 7.30
L23	8.05	4.19	1.94, 1.94	γCH 1.37; δH_3 0.92, 0.87
L24	7.59	3.49	1.31, 1.31	γCH 1.42; δH_3 0.14, 0.05
T25	7.90	4.29	4.51	γCH_3 1.35
V26	8.79	3.89	2.30	γCH_3 1.13, 1.02
K27	7.76	4.20	1.44, 1.40	γCH_2 1.72, 0.33; δH_2 1.33, 1.18; εH_2 2.61, 2.61
K28	8.06	4.04	2.18, 2.02	γCH_2 1.45, 1.52; δH_2 1.98, 1.56; εH_2 3.10, 3.10
Q29	8.31	5.02	2.31, 2.31	γCH_2 1.81, 1.81
P30		4.65	2.43, 2.14	γCH_2 2.82, 2.66; δH_2 4.16, 4.06
F31	8.42	5.66	*3.09, 2.74*	δH 7.35; εH 7.21; ξH 7.29
E32	8.71	4.62	2.12, 2.15	γCH_2 2.42, 2.42
F33	9.29	5.53	*3.39, 2.98*	δH 7.19; εH 7.03; ξH 6.95
I34	9.01	3.91	1.00	γCH_2 0.90, 0.19; γCH_3 0.65; δCH_3 0.76
N35	8.47	4.90	3.11, 2.84	
I36	7.91	4.09	2.00	γCH_2 1.02 ; γCH_3 0.83; δCH_3 0.95
M37	7.97	5.44	1.91, 1.73	γCH_2 2.42, 2.42; εCH_3 1.35
P38		4.29	2.41, 2.12	γCH_2 2.04, 2.03; δH_2 4.19, 3.40
E39	7.35	4.58	1.84, 2.08	γCH_2 2.26, 2.37
K40	8.93	3.73	1.83, 1.83	γCH_2 1.51, 1.59; δH_2 1.55,1.50; εH_2 3.11, 3.11
G41	8.11	4.20, 3.67		
V42	8.09	4.30	1.91	γCH_3 0.96, 0.83
F43	8.55	5.13	3.00, 3.00	δH 7.43; εH 7.33; ξH 6.98

D44	8.43	4.91	2.42, 2.42	
D45	8.37	4.40	2.82, 2.70	
E46	8.43	4.22	2.24, 2.24	$\gamma\mathrm{CH_2}$ 2.55, 2.43,
K47	7.87	4.24	1.94, 2.18	$\gamma\mathrm{CH_2}$ 1.55, 1.04; $\delta\mathrm{H_2}$ 1.71,1.71 $\varepsilon\mathrm{H_2}$, 3.09,3.10
I48	8.24	3.69	2.02	$\gamma\mathrm{CH_2}$ 1.59, 1.26; $\gamma\mathrm{CH_3}$ 0.77; $\delta\mathrm{CH_3}$ 0.75
A49	8.10	4.03	1.53	
E50	7.78	4.00	2.29, 2.16	$\gamma\mathrm{CH_2}$ 2.44, 2.44
L51	7.84	3.48	*1.02, 1.89*	γCH 1.18; $\delta\mathrm{H_3}$ 0.57, 0.38
L52	8.79	3.74	1.62, 1.62	γCH 1.85; $\delta\mathrm{H_3}$ 0.62, 0.51
T53	8.19	4.06	4.28	$\gamma\mathrm{CH_3}$ 1.26
K54	7.81	4.04	1.74, 1.47	$\gamma\mathrm{CH_2}$ 1.24, 0.96; $\varepsilon\mathrm{H_2}$ 2.61, 2.54
L55	7.90	4.23	1.70, 1.49	γCH 1.80; $\delta\mathrm{H_3}$ 0.66, 0.36
G56	7.97	3.98, 3.98		
R57	8.08	4.65	1.89, 1.71	$\gamma\mathrm{CH_2}$1.66, 1.22; $\delta\mathrm{H_2}$ 3.11, 3.11; εH 7.62
D58	8.63	4.69	2.83, 2.74	
T59	7.43	4.63	4.27	$\gamma\mathrm{CH_3}$ 1.22
Q60	8.93	4.24	2.16, 1.98	$\gamma\mathrm{CH_2}$ 2.06, 2.06; $\delta\mathrm{NH_2}$ 7.34, 5.81
I61	8.15	3.88	1.83	$\gamma\mathrm{CH_2}$ 1.23, 0.91; $\gamma\mathrm{CH_3}$ 1.15; $\delta\mathrm{CH_3}$ 1.56
G62	9.05	4.20, 3.78		
L63	7.79	4.42	1.81, 1.31	γCH 1.48; $\delta\mathrm{H_3}$ 0.80, 0.69
T64	8.32	4.77	4.46	$\gamma\mathrm{CH_3}$ 1.38
M65	8.73	5.24	2.52, 2.08	$\gamma\mathrm{CH_2}$ 2.24, 2.29; $\varepsilon\mathrm{CH_3}$ 1.81
P66		5.36	2.43, 2.13	$\gamma\mathrm{CH_2}$ 1.84, 1.69; $\delta\mathrm{H_2}$ 4.22, 3.45
Q67	8.62	4.90	2.38, 2.38	$\gamma\mathrm{CH_2}$ 3.02, 3.02
V68	9.39	5.41	1.92	$\gamma\mathrm{CH_3}$ 1.18, 1.04
F69	9.78	5.22	*3.13, 3.04*	δH 7.50; εH 7.23; ξH 6.83
A70	9.20	3.97	1.38	
P71		4.21	2.25, 1.85	$\gamma\mathrm{CH_2}$ 1.55, 1.23; $\delta\mathrm{H_2}$ 3.27, 3.27
D72	7.66	4.54	3.15, 2.67	
G73	8.59	4.52, 3.78		
S74	8.46	4.43	3.96, 3.96	
H75	9.36	4.16	3.30, 3.30	δ^2H 5.98; δ^1H 8.66
I76	8.21	3.89	1.18	$\gamma\mathrm{CH_2}$ 0.81, 1.37; $\gamma\mathrm{CH_3}$ 0.49; $\delta\mathrm{CH_3}$ 0.19
G77	6.94	4.37, 3.20		
G78	8.21	4.40, 4.32		
F79	9.15	4.09	*3.29, 2.89*	δH 7.35; εH 7.35; ξH 6.98
D80	8.74	3.95	2.60, 2.44	
Q81	7.15	3.98	2.00, 1.83	$\gamma\mathrm{CH_2}$ 2.74, 2.41; $\delta\mathrm{NH_2}$ 7.89, 7.17
L82	8.18	3.92	*2.07, 1.89*	γCH 1.75; $\delta\mathrm{H_3}$ 0.98, 0.81
R83	8.07	3.86	1.69, 1.44	$\gamma\mathrm{CH_2}$ 131, 1.31; $\delta\mathrm{H_2}$ 3.17, 3.05; εH 7.08
E84	7.24	4.09	2.05, 2.04	$\gamma\mathrm{CH_2}$ 2.48, 2.43
Y85	7.83	3.99	*2.96, 3.07*	δH 6.69; εH 6.31
F86	7.07	4.62	*2.84, 3.52*	δH 7.54; εH 7.20; ξH 6.80
K87	7.61	4.08	1.86, 1.73	$\gamma\mathrm{CH_2}$ 1.48, 1.48; $\delta\mathrm{H_2}$ 1.60, 1.64; $\varepsilon\mathrm{H_2}$ 3.03, 3.03

* Stereospecific assignments of methylene protons are printed in italics and underlined. The first number corresponds to β^2 proton and the second to the β^3 proton.

Table 4.2. ^{1}H chemical shift* assignments for reduced bacteriophage T4 Grx

residue	NH	αH	βH	others
M1		4.40	2.20,2.33	γCH$_2$ 2.66, 2.54; εCH$_3$ 2.02
F2	8.48	4.67	3.25, 2.83	δH 7.40; εH 7.22; ξH 7.48
K3	8.71	5.24	1.92, 1.58	γCH$_2$ 1.48, 1.35; δH$_2$ 1.99, 1.70; εH$_2$ 2.91, 2.91
V4	9.01	4.83	2.30	γCH$_3$ 0.90, 0.85
Y5	9.48	5.20	*2.64, 3.20*	δH 7.11; εH 6.54
G6	8.47	4.24, 4.00		
Y7	8.25	5.20	*2.94, 3.68*	δH 7.32; εH 6.83
D8	8.70	4.74	2.91, 2.84	
S9	7.36	4.11	4.57, 4.36	
N10	8.39	4.49	3.20, 3.03	
I11	8.13	4.08	2.08	γCH$_2$ 1.20, 0.93; γCH$_3$ 1.01; δCH$_3$ 0.85
H12	8.77	4.50	*2.79, 2.73*	δ^{2}H 8.34 ; δ^{1}H 5.81
K13	7.99	4.20	1.58, 1.65	γCH$_2$ 1.72, 1.73; δH$_2$ 2.01, 2.01; εH$_2$ 3.14, 3.14
C14	8.23	4.82	2.74, 2.76	
V15	8.00	3.82	1.88	γCH$_3$ 0.92, 0.86
Y16	9.05	4.08	3.27, 2.92	δH 7.24 ; εH 7.05
C17	8.25	4.70	3.21, 2.96	
D18	8.15	3.95	2.80, 2.66	
N19	8.58	4.39	2.16, 2.16	δNH$_2$ 7.04, 5.93
A20	7.95	4.16	1.53	
K21	7.68	3.30	1.52, 1.52	γCH$_2$ 1.87, 1.22; δH$_2$ 1.94, 1.94; εH$_2$
R22	8.10	4.06	1.91, 1.91	γCH$_2$ 1.67, 1.74; δH$_2$ 3.24, 3.24; εH 7.32
L23	8.09	4.08	1.79, 1.79	γCH 1.38; δH$_3$ 0.95, 0.91
L24	7.52	3.45	1.26, 1.26	γCH 1.43; δH$_3$ 0.16, 0.08
T25	7.85	4.23	4.45	γCH$_3$ 1.37
V26	8.75	3.87	2.24	γCH$_3$ 1.12, 1.00
K27	7.71	4.18	1.44, 1.37	γCH$_2$ 1.70, 0.32; δH$_2$ 1.32, 1.17; εH$_2$ 2.62, 2.62
K28	7.98	3.98	2.10, 1.99	γCH$_2$ 1.42, 1.51; δH$_2$ 1.95,1.96; εH$_2$ 3.12, 2.99
Q29	8.27	4.98	2.26, 2.26	γCH$_2$ 1.83,1.83
P30		4.67	2.40, 2.16	γCH$_2$ 2.81, 2.65; δH$_2$ 4.06, 3.97
F31	8.16	5.67	*3.10, 2.73*	δH 7.36; εH 7.24; ξH 7.28
E32	8.84	4.54	1.98, 1.89	γCH$_2$ 2.25, 2.25
F33	9.31	5.42	*3.28, 2.92*	δH 7.21; εH 7.04; ξH 6.93
I34	9.03	3.89	1.05	γCH$_2$ 0.72, 0.22; γCH$_3$ 0.66; δCH$_3$ 082
N35	8.43	5.03	3.08, 2.87	
I36	8.51	4.34	2.00	γCH$_2$ 1.02 ; γCH$_3$ 0.84; δCH$_3$ 0.72
M37	7.92	5.43	1.91, 1.73	γCH$_2$ 2.42, 2.36; εCH$_3$ 1.35
P38		4.30	2.40, 2.13	γCH$_2$ 2.08, 1.97; δH$_2$ 4.16, 3.36
E39	7.31	4.48	1.95, 1.79	γCH$_2$ 2.20, 2.17
K40	8.92	3.72	1.85, 1.85	γCH$_2$ 1.53, 1.61; δH$_2$ 1.59, 1.52; εH$_2$ 3.12, 3.12
G41	8.07	4.17, 3.64		
V42	8.05	4.35	1.85	γCH$_3$ 0.92, 0.77
F43	8.38	5.13	2.96, 2.96	δH 7.43; εH 7.33; ξH 6.98

D44	8.44	4.82	2.47, 2.47	
D45	8.35	4.42	2.84, 2.72	
E46	8.33	4.18	2.22, 2.16	γCH_2 2.55, 2.39
K47	7.88	4.23	1.90, 2.14	γCH_2 1.52, 1.03; δH_2 1.71,1.71 εH_2 , 3.07, 3.07
I48	8.26	3.68	2.04	γCH_2 1.57, 1.24; γCH_3 0.73; δCH_3 0.76
A49	8.05	3.98	1.50	
E50	7.72	3.96	2.02, 2.02	γCH_2 2.30, 2.30
L51	7.79	3.45	*0.98, 1.83*	γCH 1.15; δH_3 0.53, 0.37
L52	8.73	3.69	1.62, 1.62	γCH 1.84; δH_3 0.62, 0.51
T53	8.09	4.03	4.35	γCH_3 1.20
K54	7.77	4.10	1.74, 1.47	γCH_2 1.24, 0.96; εH_2 2.61, 2.54
L55	7.90	4.16	1.70, 1.49	γCH 1.80; δH_3 0.66, 0.36
G56	7.97	3.98, 3.98		
R57	8.08	4.65	1.89, 1.71	$\gamma CH_2$1.66, 1.22; δH_2 3.11, 3.11; εH 7.62
D58	8.63	4.69	2.83, 2.74	
T59	7.43	4.63	4.27	γCH_3 1.22
Q60	8.44	4.24	2.16, 1.98	γCH_2 2.06, 2.06; δNH_2 7.34, 5.81
I61	8.15	3.84	1.82	γCH_2 1.23, 0.91; γCH_3 1.24; δCH_3 1.56
G62	9.05	4.23, 3.74		
L63	7.69	4.33	1.73, 1.46	γCH 1.26; δH_3 0.89, 0.69
T64	8.27	4.50	4.15	γCH_3 1.36
M65	8.38	5.13	2.24, 2.15	γCH_2 2.48, 2.65; εCH_3 1.81
P66		5.28	2.43, 2.13	γCH_2 1.99, 1.90; δH_2 4.07, 3.60
Q67	8.51	4.77	2.38, 2.38	γCH_2 3.02, 3.02
V68	9.18	5.36	1.87	γCH_3 1.10, 1.10
F69	9.69	5.17	*3.09, 3.00*	δH 7.50; εH 7.23; ξH 6.83
A70	9.20	3.97	1.35	
P71		4.21	2.25, 1.85	γCH_2 1.55, 1.23; δH_2 4.04, 3.47
D72	7.48	4.49	3.09, 2.60	
G73	8.52	4.45,3.71		
S74	8.35	4.37	3.93, 3.93	
H75	9.66	4.29	3.60, 3.60	$\delta^2 H$ 5.75; $\delta^1 H$ 8.28
I76	8.20	3.89	1.63	γCH_2 0.70, 1.14; γCH_3 0.44; δCH_3 0.19
G77	6.69	4.25, 3.18		
G78	8.07	4.05, 3.88		
F79	9.15	4.17	*3.27, 3.08*	δH 7.35; εH 7.35; ξH 6.98
D80	8.68	3.93	2.60, 2.44	
Q81	7.04	3.95	2.00, 1.83	γCH_2 2.71, 2.38; δNH_2 7.89, 7.17
L82	8.11	3.90	2.07, 1.89	γCH 1.75; δH_3 0.98, 0.81
R83	7.96	3.81	1.58, 1.44	γCH_2 132, 1.32; δH_2 3.17, 3.05; εH 7.08
E84	7.16	4.05	2.10, 1.97	γCH_2 2.36, 2.28
Y85	7.76	3.99	*2.96, 3.23*	δH 6.69; εH 6.31
F86	7.04	4.57	*2.78, 3.47*	δH 7.54; εH 7.20; ξH 6.80
K87	7.46	4.02	1.83, 1.67	γCH_2 1.48, 1.44; δH_2 1.68, 1.68; εH_2 3.03, 3.03

***Stereospecific assignments of methylene protons are printed in italics and underlined. The first number corresponds to β^2 proton and the second to the β^3 proton.**

Table 4.3. Structural and stereochemical statistics for oxidized and reduced T4 Grx[a]

A. Structural statistics		
A1. RMS deviation	Oxidized (SA)	Reduced (SA)
rmsd from exptl distance restraints (Å)	0.04±0.01	0.04±0.01
rmsd from exptl dihedral restraints (deg)	0.23±0.04	0.25±0.04
rmsd from $^3J_{HNN\alpha}$ restraints (Hz)	0.15±0.06	0.17±0.07
rmsd from ^{1}H chem. shift restraints (ppm)	0.31±0.10	0.28±0.12
A2. Deviations from idealized covalent geometry		
bonds (Å)	0.0036±0.0011	0.0034±0.0012
angles (deg)	0.51±0.02	0.49±0.05
impropers (deg)	0.43±0.05	0.38±0.04
A3. Procheck[b] assessment		
overall G-Factor	0.11±0.03	0.12±0.03
ϕ/ψ in the most favorable region (%)	82.1±2.4	83.2.±2.2
ϕ/ψ in additionally allowed region (%)	16.5±3.0	15.1±2.9
ϕ/ψ in generously allowed region (%)	1.4±1.7	1.6±1.8
ϕ/ψ in disallowed region (%)	0	0
H-bond energy (kcalmol^{-1})	0.78 ± 0.08	0.81 ± 0.10
Van der Waals energy (kcalmol^{-1})	-349.4 ± 38.4	-351.1 ± 33.4
number of bad contacts/100 residues	1.9±1.2	2.1 ± 1.4

B. Atomic coordinate rms deviation (Å)[c]				
	Oxidized		Reduced	
	backbone	heavy	backbone	heavy
<SA> vs (SA)$_{mean}$	0.71 ± 0.10	1.27 ± 0.11	0.70 ± 0.11	1.28 ± 0.12
<SA> vs X-ray[d]	1.40 ± 0.13	2.34 ± 0.20	1.39 ± 0.16	2.36 ± 0.21
(SA)$_{mean}$ vs X-ray	1.21	1.96	1.23	2.01

[a] The notation of the structures is as follows: <SA> are the final 30 NMR simulated annealing structures; (SA)$_{mean}$ is the mean structure obtained by averaging the coordinates of the individual SA structures best fit to each other.

[b] Analyzed using PROCHECK-NMR version 3.4.4 (Laskowski et al., 1996).

[c] Analyzed using MolMol version 4.0 (Koradi et al., 1996).

[d] X-ray is the 1.45 Å resolution X-ray structure of T4 Grx of Eklund et al. (1992).

Table 4.4. Amide ^{1}H half height linewidth $(\Delta\nu_{1/2})^a$ for oxidized and reduced T4 Grx.

$(\Delta\nu_{1/2})$ (Hz)			$(\Delta\nu_{1/2})$ (Hz)		
Residue	Oxidized	Reduced	Residue	Oxidized	Reduced
F2	14.57	15.74	F43	14.23	14.39
K3	20.32	21.74	K47	16.66	16.16
V4	21.15	22.34	I48	12.62	14.94
Y5	20.43	21.75	E50	12.18	12.49
H12	14.09	13.93	L51	11.07	12.33
K21	12.02	14.84	L52	11.15	13.38
L24	12.78	14.21	T53	11.84	11.72
T25	11.53	12.78	L63	12.85	14.67
V26	11.82	12.69	V68	20.05	21.14
K27	16.48	17.20	F69	21.56	22.95
F31	15.88	16.53	A70	13.93	13.68
E32	21.50	24.61	D80	10.29	12.89
F33	20.81	22.12	Q81	15.17	15.82
I34	19.50	20.36	E84	12.90	13.10
M37	19.76	21.68	Y85	12.67	14.68
E39	15.41	16.85	F86	15.72	16.52

[a] From 600 MHz TOCSY spectra processed with the sine-bell square weight function (sb=-0.100, sbs=-0.066).

Table 4.5. $^3J_{HNN\alpha}$ coupling constants for oxidized and reduced T4 Grx[a]

$^3J_{HNN\alpha}$				$^3J_{HNN\alpha}$			
Residue	X-ray[b]	Oxidized	Reduced	Residue	X-ray[b]	Oxidized	Reduced
F2	5.89	6.10	6.16	D44	7.34	5.87	6.10
K3	9.40	10.4	10.1	E46	4.76	3.70	3.90
V4	9.28	9.40	10.4	K47	6.48	6.70	7.50
Y5	8.70	8.96	8.90	I48	5.33	5.26	4.70
Y7	8.81	9.61	10.4	A49	4.88	4.60	4.70
N10	5.53	4.13	4.50	E50	4.14	5.00	5.45
I11	8.95	8.23	8.40	L51	4.40	4.96	4.96
H12	**8.35**	**4.60**	**5.71**	L52	4.03	4.30	4.23
K13	4.01	4.42	4.10	T53	4.27	3.19	3.85
V15	5.57	5.53	5.60	K54	4.55	4.89	4.70
Y16	5.90	5.78	5.50	L55	6.55	7.02	6.81
C17	5.15	5.20	5.34	R57	9.32	8.15	8.31
N19	4.27	4.63	4.51	D58	7.18	7.05	7.20
A20	3.76	4.00	3.90	T59	8.03	7.57	7.70
K21	5.84	4.85	4.05	Q60	6.66	7.12	7.51
R22	5.01	3.48	3.61	I61	4.51	4.36	4.95
L23	4.18	3.84	4.10	L63	4.09	5.62	5.31
L24	5.10	4.53	4.42	T64	9.01	7.91	7.32
T25	5.57	4.80	4.81	M65	9.39	9.82	9.60
V26	4.81	4.69	4.86	Q67	9.36	9.30	9.70
K27	7.18	6.48	6.72	**V68**	**9.40**	**7.76**	**7.81**
Q29	8.88	8.60	9.20	**F69**	**9.40**	**7.73**	**7.63**
F31	7.92	7.10	7.41	A70	4.90	4.81	4.95
E32	9.30	9.96	9.10	D72	7.10	5.73	5.80
F33	8.32	8.47	8.41	S74	5.05	5.02	4.82
I34	9.29	9.17	9.34	**H75**	**6.29**	**4.14**	**4.30**
N35	7.85	6.35	7.20	I76	8.56	7.53	7.80
I36	**6.92**	**7.30**	**5.70**	F79	4.55	5.10	5.23
M37	8.31	8.54	8.24	D80	5.13	3.86	3.42
E39	7.43	6.64	6.51	Q81	5.21	5.97	6.24
K40	3.45	3.60	3.73	E84	5.10	4.74	5.24
V42	9.17	9.30	9.40	Y85	5.53	4.07	3.81
F43	7.95	6.06	6.12	F86	8.71	6.76	7.06

[a] Strongest differences between X-ray and observed $^3J_{HNN\alpha}$ are highlighted.

[b] 1.45 Å resolution X-ray structure of T4 Grx of Eklund et al. (1992).

[c] The molecular weights used in the calculation of $^3J_{HNN\alpha}$ were 20100 and 10050 for oxidized and reduced T4 Grx respectively.

Table 4.6. X-ray and NMR structures of solved thioredoxins and glutaredoxins

Protein name	Redox status	Resolution	Reference (PDB code)
E. coli Trx	Oxidized	2.8 Å X-ray	Holmgren et al., 1975 (N/A)
	Oxidized	1.6 Å X-ray	Katti et al., 1990 (2TRX)
	Oxidized	NMR	Jeng et al., 1994 (1XOA)
	Reduced	NMR	Dyson et al., 1990 (N/A)
	Reduced	NMR	Jeng et al., 1994 (1XOB)
E. coli Grx	Oxidized	NMR	Xia et al., 1992 (1EGO)
	Reduced	NMR	Sodano et al., 1991 (1EGR)
	Mixed disufide	NMR	Bushweller et al., 1994 (1GRX)
Human Trx	Oxidized	2.1 Å X-ray	Weichsel et al., 1996 (1ERU)
	Reduced	1.7 Å X-ray	Weichsel et al., 1996 (1ERT)
	Reduced	NMR	Forman-Kay et al., 1991 (4TRX)
	Reduced	NMR	Qin et al., 1994 (N/A)
Human Grx	Reduced	NMR	Sun et al., 1998 (1JHB)
Calf thymus Grx	Oxidized	2.2 Å X-ray	Katti et al., 1995 (1KTE)
T4 Grx	Oxidized (wt)	2.8 Å X-ray	Soderberg et al., 1978 (OTT4)
	Oxidized (wt)	2.0 Å X-ray	Eklund et al., 1992 (1AAZ)
	Oxidized (wt)	NMR	this paper (N/A)
	Oxidized (mut)	1.45 Å X-ray	Eklund et al., 1992 (1ABA)
	Reduced (mut)	2.4 Å X-ray	Ingelman et al., 1995 (N/A)
	Reduced (wt)	NMR	this paper (N/A)

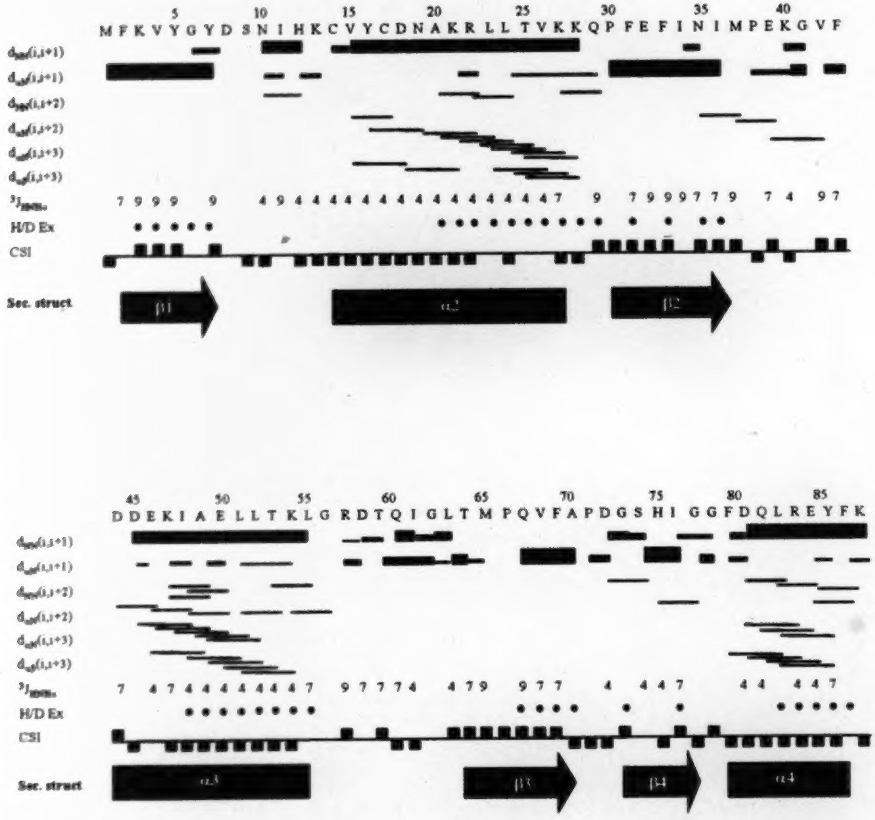
Table 4.7. Atomic accessible surface area for cysteines in the active site area (Å^2)[a] versus measured pKa values.

Molecule	N (Å^2)	CA(Å^2)	CB(Å^2)	SG(Å^2)	pKa
E. coli. Trx (PDB code: 1OXB)					
Cys 32	0.0±0.0	0.0±0.0	0.0±0.0	11.1± 1.8	7.1
Cys 35	0.0±0.0	0.0±0.0	0.2±0.2	0.0±0.0	9.9
T4 Grx					
Cys 14[b]	3.2±1.3	0.0±0.2	0.5±0.7	11.6±4.5	6.8
Cys 14[c]	0.4	2.3	0.4	10.2	
Cys 17[b]	0.3±0.3	0.0±0.0	0.3±0.4	1.10±1.3	9.5
Cys 17[c]	0.0	0.4	1.1	0.0	
Pig liver Grx (PDB code: 1KET)					
Cys 22	6.8	1.1	2.3	11.4	2.5
Cys 25	0.0	0.4	4.6	0.0	N/A

[a] The atomic accessible surface area is calculated using MoMol version 4.0 (Koradi et al., 1996).

[b] The NMR structures (from this study).

[c] The X-ray structure, PDB code 1ABA (Eklund et al., 1992).



M F K V Y G Y D S N I H K C V Y C D N A K R L L T V K K Q P F E F I N I M P E K G V F
dNN(i,i+1)
dαN(i,i+1)
dNN(i,i+2)
dαN(i,i+2)
dαN(i,i+3)
dαβ(i,i+3)
3JHNNα
H/D Ex
CSI
Sec. struct
β1
α2
β2
D D E K I A E L L T K L G R D T Q I G L T M P Q V F A P D G S H I G G F D Q L R E Y F K
α3
β3
β4
α4

Figure 4.1. Schematic representation of the secondary structure determination of T4 glutaredoxin with summary of sequential and medium region (i, i+3, n≤3) NOE, $^{3}J_{HNN\alpha}$ coupling constants (4 for $^{3}J_{HNN\alpha} \leq 6$ Hz; 7 for 6 Hz< $^{3}J_{HNN\alpha}$ <8 Hz; 9 for $^{3}J_{HNN\alpha} \geq 9$ Hz), slow amide proton exchange rate (marked with a •), chemical shift index (CSI) (upward and downward ■ represent upfield and downfield α-proton chemical shifts respectively). On the last line of each segment, helices and β-strands as determined from above data are indicated as rectangles and arrows respectively.

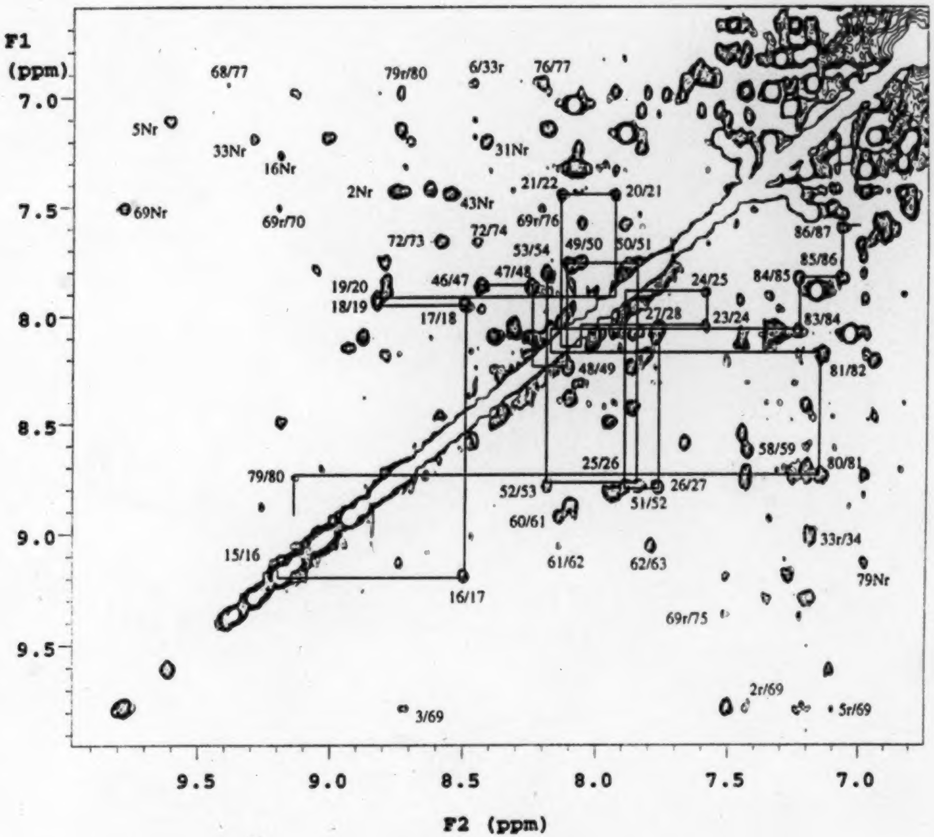


Figure 4.2 HN and aromatic region of the phase-sensitive 500-MHz NOESY spectrum of oxidized T4 Grx. "N" denotes amide protons, "r" denotes aromatic ring protons and "Nr" denotes the intra-residue amide/ring proton intercation.

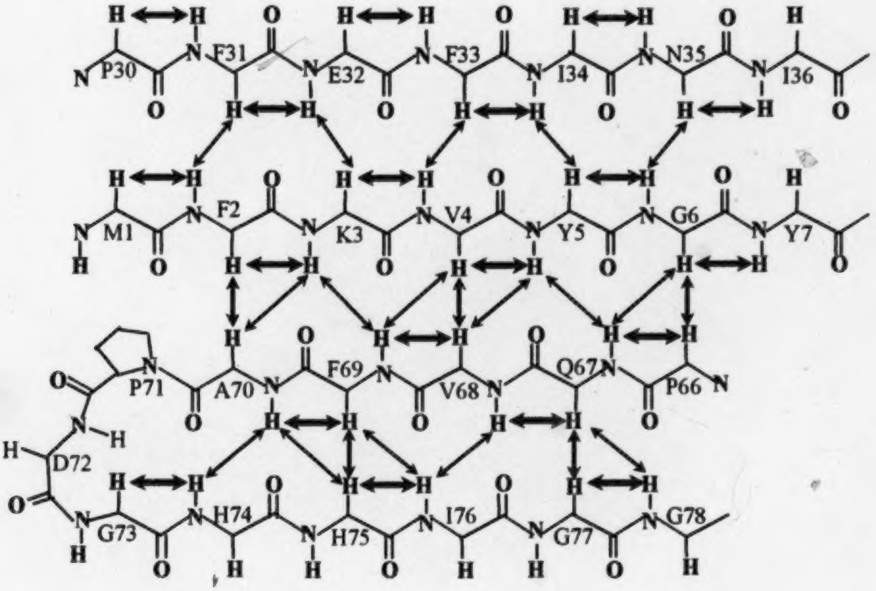
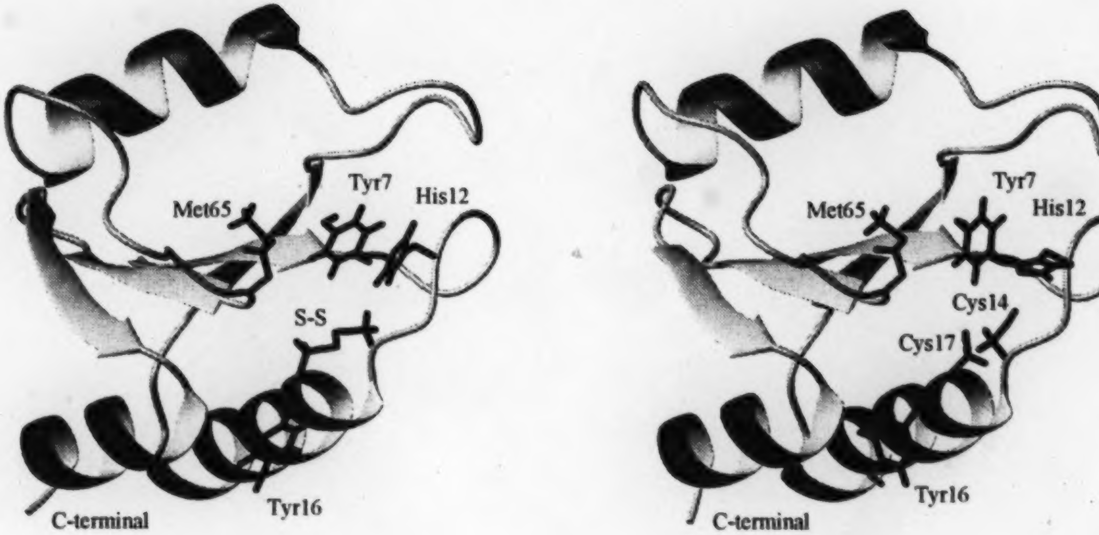

P30
F31
E32
F33
I34
N35
I36
M1
F2
K3
V4
Y5
G6
Y7
P71
A70
F69
V68
Q67
P66
D72
G73
H74
H75
I76
G77
G78

Figure 4.3. Schematic representation of four-stranded β-sheet and NOE connectivity's observed.


Met65
Tyr7
His12
S-S
C-terminal
Tyr16
Met65
Tyr7
His12
Cys14
Cys17
C-terminal
Tyr16

(a) Oxidized T4 glutaredoxin

(a) Reduced T4 glutaredoxin

Figure 4.4. Ribbon representations restrained minimized mean structure of (a) oxidized and (b) reduced T4 glutaredoxin. The side chains of several residues around the redox-active site region are displayed and labeled.

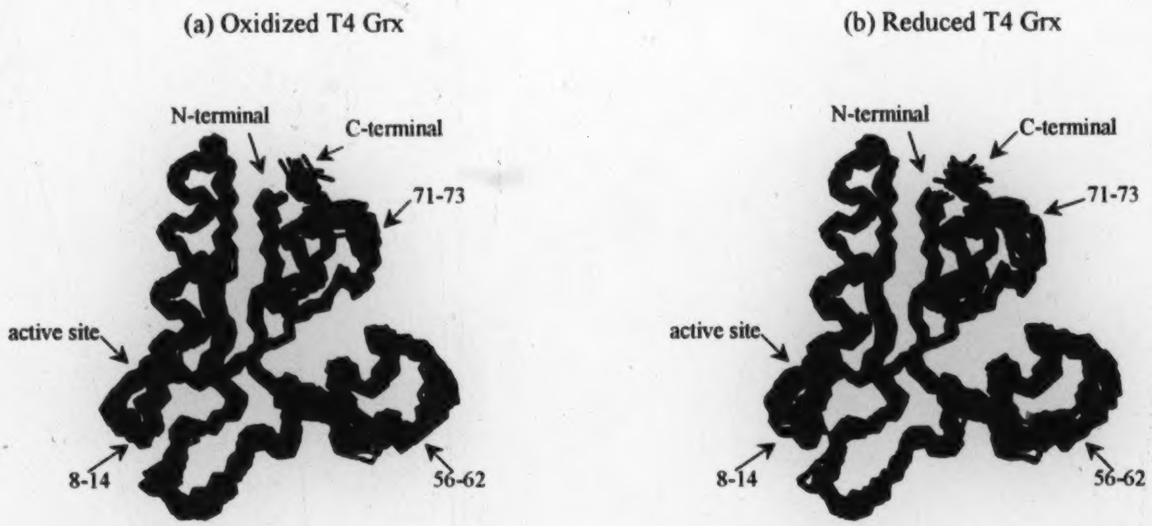

(a) Oxidized T4 Grx
N-terminal
C-terminal
71-73
active site
8-14
56-62
(b) Reduced T4 Grx
N-terminal
C-terminal
71-73
active site
8-14
56-62

Figure 4.5. 30 superimposed backbone atom traces for (a) oxidized and (b) reduced T4 glutaredoxin

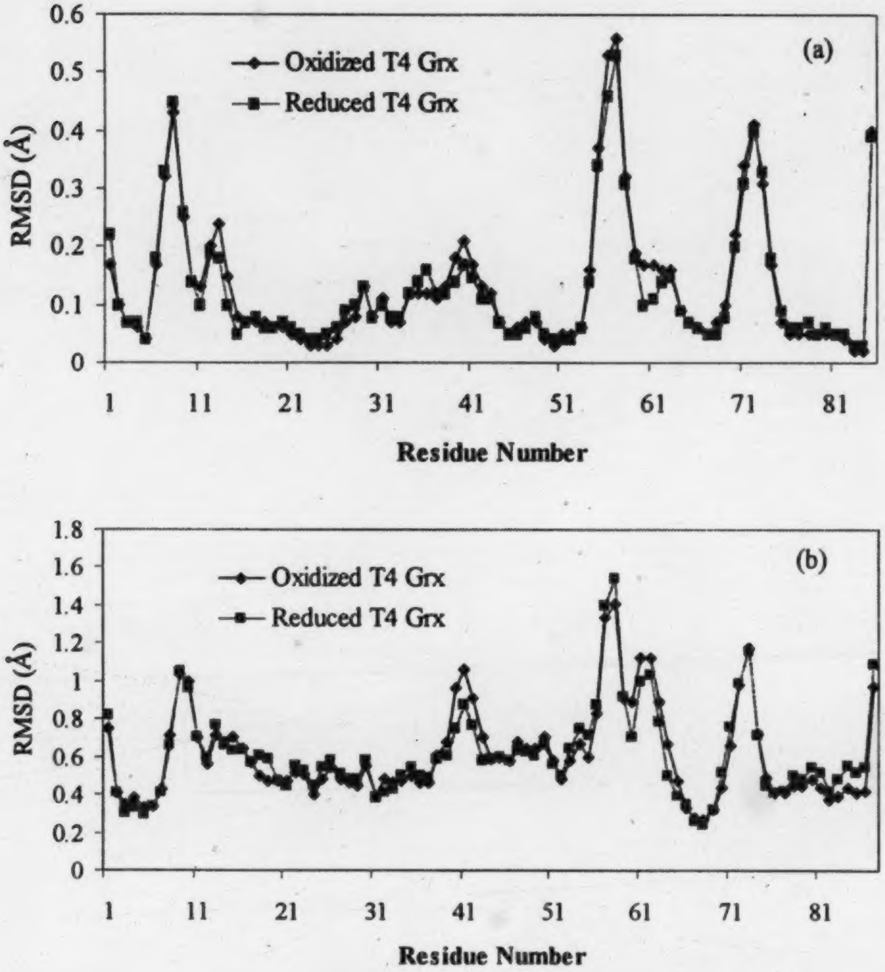


Figure 4.6. Atomic rms distribution for the backbone superposition of the final set of structures about their mean coordinates, (a) local rmsd values for the backbone superposition of all tripeptides along the sequence are plotted at the positions of the central resides (b) overall rmsd values for the complete backbone superpostion.

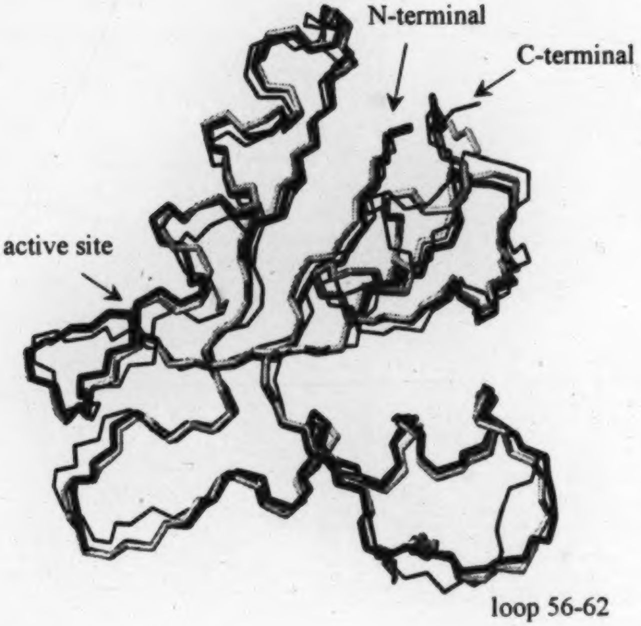

N-terminal
C-terminal
active site
loop 56-62

Figure 4.7. Best fit superposition of the backbone atoms of the energy minimized mean NMR structures of oxidized (bold black), reduced (bold grey) and 1.45 Å X-ray (thin black) structure (Eklund et al., 1992) of T4 Grx.

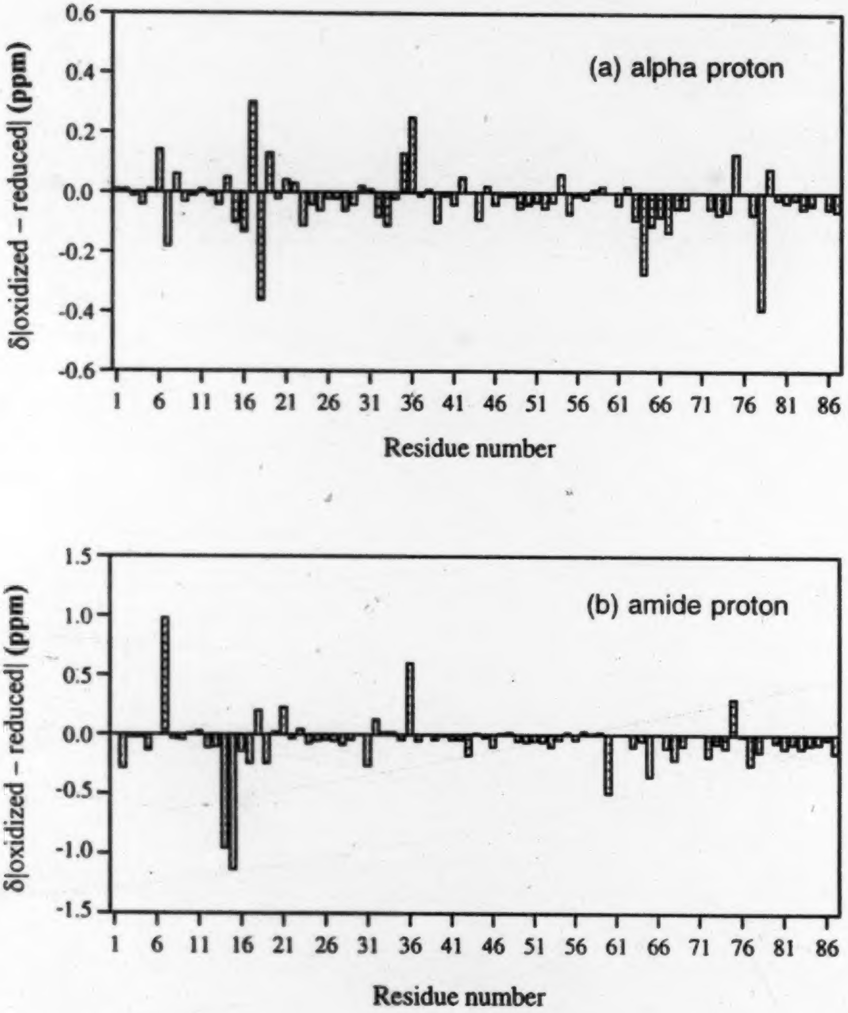


Figure 4.8. α-proton (a) and amide proton (b) chemical shift differences between oxidized and reduced T4 Grx.

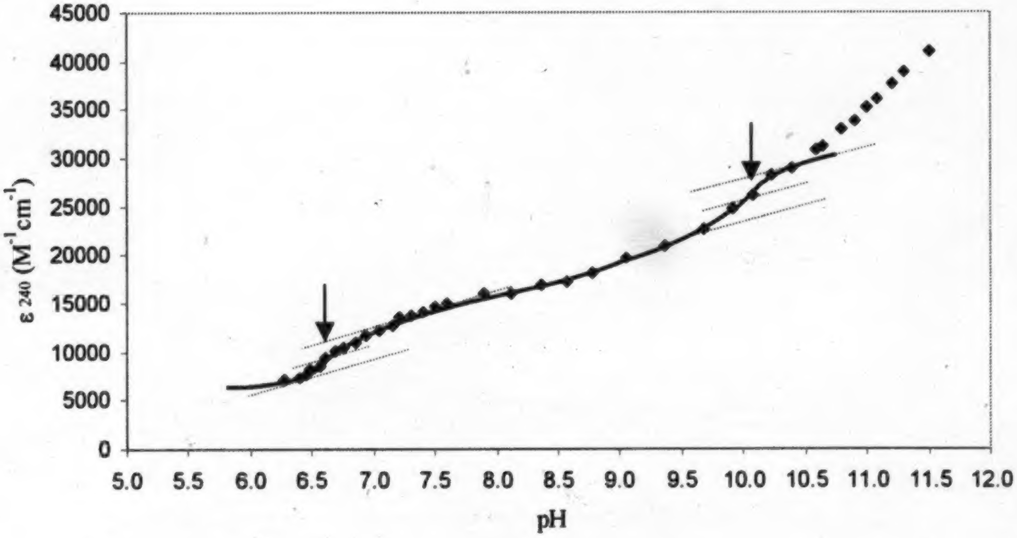
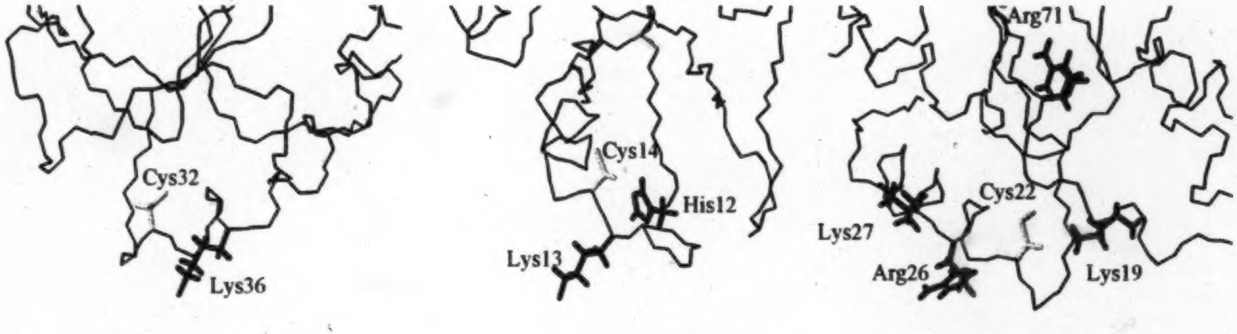


Figure 4.9. Measurement of thiol ionization in T4 Grx by ultraviolet absorbance.


Cys32
Lys36
Cys14
His12
Lys13
Arg71
Cys22
Lys27
Arg26
Lys19

(a) E. coli thioredoxin (b) T4 glutaredoxin (c) Pig liver glutaredoxin

Figure 4.10. Stereo-view of the exposed cysteines and the surrounding charged residues for (a) E. coli thioredoxin, (b) T4 glutaredoxin, and (c) pig liver glutaredoxin.

Chapter 5*

The Solution Structure of Carnobacteriocin B2: Implications for Structure-Activity Relationships Among Type IIa Bacteriocins

5.1. Introduction

Over the past two decades, dozens of antimicrobial peptides from a wide variety of plants, mammals, and micro-organisms have been identified, isolated and characterized (Cammue et al., 1994; Boman, 1995; Nissen-Meyer and Nes 1997). Many of these antimicrobial peptides are lethal to bacteria, fungi and even viruses (Jacob and Zasloff, 1994). Despite their broad spectrum of antimicrobial activities, the majority of these peptides display little toxicity towards normal mammalian cells. Because of their unusual specificity, it has been proposed that naturally occurring antimicrobial peptides could eventually provide a viable alternative to conventional antibiotics, many of which are being rendered ineffective due to the emergence of new and dangerously resistant bacterial strains (Nikaido, 1994; Davies, 1994).

One class of antimicrobial peptides that is attracting increasing attention are the so-called bacteriocins (Bennik et al., 1998). Bacteriocins are bacterially derived, ribosomally synthesized antimicrobial peptides. Those produced by Gram-positive organisms, particularly lactic acid bacteria (LAB), have exhibited considerable potential as non-toxic preservatives in the food and feed industry (Vanderbergh, 1993; Deles-Broughton et al., 1996). Based on their probable structures and mechanisms of action, LAB bacteriocins have been classified into four different classes (types I, II, III and IV). Type I and II LAB bacteriocins are small, heat-stable peptides exhibiting a cationic and hydrophobic character. Their antimicrobial activity generally seems to be related to their ability to form pores in the cytoplasmic membrane of target cells. Type III LAB bacteriocins are typically large, heat-labile proteins whose mechanism of action still remains unknown; while Type IV bacteriocins are generally more complex molecules composed of at least one protein and one or more non-protein moieties required for activity (Klaenhammer, 1993).

Carnobacteriocin B2 (CbnB2) is a cationic LAB bacteriocin that occurs in the culture supernatant of *Carnobacterium piscicola* LV17B, a Lactobacilli originally

---

*This chapter has been submitted to *Biochemistry* by Wang, Y. et al., 1999 isolated from refrigerated meats (Quadri et al., 1994, 1995 and 1997). It is believed that CbnB2 targets and disrupts the cytoplasmic membrane of sensitive cells, causing dissipation of their proton motive force and concomitant leakage of their intracellular components (Quadri et al., 1997). CbnB2's spectrum of antimicrobial activity includes many LAB as well as strains of potentially pathogenic Enterococcal and Listerial species. CbnB2 is ribosomally synthesized as a 66 residue precursor peptide, called precarnobacteriocin B2. Post-translational cleavage at a Gly^{-2}-Gly^{-1} site to remove an 18-amino acid leader sequence from the N-terminus yields the final, fully active 48 residue peptide (i.e. CbnB2).

CbnB2 belongs to the second subclass (Type IIa) LAB bacteriocins, which is characterized by a well-conserved heptapeptide motif: Tyr-Gly-Asn-Gly-Val-Xaa-Cys and a strongly conserved disulfide bond located near the N-terminus. This YGNGVXC motif has been suggested to form part of a recognition sequence for a putative membrane-bound protein receptor (Fimland et al., 1996; Fleury et al., 1996). To date, the only member in this subclass whose three-dimensional structure has been solved is Leucocin A (LeuA), a 37 amino acid antimicrobial peptide isolated from *Leuconostoc gelidum* UAL187 (Hastings et al., 1991; Van and Stiles, 1995). NMR studies show that LeuA adopts well-defined tertiary structure which is composed of a three-stranded antiparallel β-sheet (residues 2 to 16) followed by an amphiphilic α-helix (residues 17 to 31) in both TFE and dodecylphosphocholine (DPC) micelles. Interestingly, the same peptide is essentially unstructured in water or aqueous DMSO (Sailer et al., 1993; Fregeau et al., 1997).

CbnB2 exhibits considerable sequence similarity to LeuA as well as to other Type IIa bacteriocins, especially near the N-terminus, where the characteristic heptapeptide motif (YGNGVXC) is highly conserved (Klaenhammer, 1993; Quadri et al., 1994). However, unlike LeuA or other Type IIa bacteriocins, CbnB2 does not appear to form a disulfide bond between its two conserved cysteines, Cys9 and Cys14 (Quadri et al., 1997). In an effort to determine the structural consequences of this "non-existent" disulfide and to further ascertain if the YGNGVXC sequence motif actually confers a common "receptor-binding" structural motif (Fleury et al., 1996), we have determined the three-dimensional solution structure of CbnB2 using NMR spectroscopy. In addition we have attempted to verify the redox status of

CbnB2's two conserved cysteines using model peptides, mass spectrometry and chemical modification experiments. By determining the true redox status of CbnB2 and by comparing the structure of CbnB2 with that of LeuA we believe it may now be possible to confirm or clarify several key hypotheses about the structure-activity relationships of this unique class of peptides.

5.2. Experimental

Peptide Synthesis, Purification and Characterization. To study the stability and sequence dependent effects of the putative CbnB2 disulfide bond, six hexapeptides were prepared. Their sequences are listed in Table 5.1. All six peptides have cysteines in positions 1 and 6, and all are acetylated at the N-terminus. Peptide #1 corresponds to the sequence found between residues 9 -14 in CbnB2, while peptide #2 corresponds to the sequence found in LeuA. The remaining four peptides are homologues of peptide #1 but with varying amino acid substitutions. A seventh peptide, spanning residues 1 - 22 of CbnB2, was also synthesized to investigate whether the redox status of CbnB2 was affected by its C terminal helix. All peptides were synthesized on a Rainin PS 3 peptide synthesizer using standard Fmoc solid phase peptide chemistry (Erickson and Merrifield, 1976; Meienhofer et al., 1979). For each peptide, the C terminal amino acid was activated with HBTU/NMM/DMF (1:1:2) and allowed to react with Wang resin (0.2 mmol). Each amino acid was assembled in turn by automated cycles of deprotection, activation and coupling. The amino acids were deprotected by reacting with 20% piperdine in DMF for 10 minutes followed by washing with DCM (3x10 ml) and DMF (5x10 ml). Each residue was activated by mixing the protected amino acid (0.8 mmole in 2.7 ml DMF) with HBTU (0.7 mmole in DMF) and a slight excess (1.6 mmole) of neat NMM for 5 minutes. The activated amino acids were condensed to the peptide resin by agitation (N_2 bubbling) for 45 minutes and then washed with DCM (3x10 ml) and DMF (5x10 ml). The N-terminal acetylation was carried out by reacting with NMM (800 μl), acetic anhydride (500 μl) and DMF (15 μl) for 15 minutes. The completed peptide was cleaved with 10 ml of a cleavage cocktail (95%TFA:5%H_2O) for two hours and then precipitated with cold ether (30 ml). Crude peptides were purified by reverse-phase HPLC on a reversed-phase C-18 21 x 250 mm Zorbax 300 SB column running a binary (H_2O:acetonitrile) gradient at a flow rate of 8 ml/min using aqueous

0.1% TFA and 0.1% TFA in MeCN as the mobile phase. The eluent was monitored at 220 nm and positive fractions were analyzed by HPLC and ES-MS to confirm their identity.

Kinetic Studies of Hexapeptide Redox Reaction Rates For the kinetic study of the six CbnB2 and LeuA hexapeptide mimics, 1 mg of each oxidized peptide was dissolved in 280 µl of a 45 mM deuterated phosphate/TFE buffer (50%), pH 5.2. A dithiothreitol (DTT) solution was prepared by dissolving 9.75 mg DTT in 1 ml of the same buffer (minus TFE). An aliquot of 50 µl of this solution, corresponding to ~7 equivalents of DTT, was added to each peptide solution. The concentration of each peptide was determined by ^{1}H NMR peak integration of the DTT resonances (methylene protons) and selected peptide resonances (N-terminal acetyl group protons and threonine γ-protons). The reduction of each peptide was monitored by 1D and 2D ^{1}H NMR and the kinetics measured by the time-dependent loss of peak intensity of resonances corresponding to β-protons of oxidized cysteine (~ 3.2 ppm). After the reduction reaction was complete, the pH was adjusted to 8.0 and the solution degassed with excess oxygen. The subsequent oxidation reaction was monitored by 1D and 2D ^{1}H NMR spectroscopy as described above.

Production and Isolation of CbnB2. CbnB2 was prepared as previously described (Quadri, et al., 1994). Briefly, 50 ml of Carnobacterium piscicola LV17B preculture was inoculated into 5 liters of medium containing (per liter): casamino acids (15 g); yeast extract (5 g); glucose (25 g); $K_2HO_43H_2O$ (1g); $MgSO_4$ (0.1 g); $MnSO_44H_2O$ (0.05 g); and Tween 80 (1ml). The culture was maintained at pH 6.2 by adding of 1 M NaOH (controlled by Chem-CadetÆ, Cole-Parmer, Chicago, IL), and gently stirred under an N_2 atmosphere for twenty-six hours at 25 °C. After centrifugation, the culture supernatant was loaded onto a 4.5 X 50 cm Superlite DAX-8 column (Sigma-Aldrich Co.) pre-equilibrated with 0.1% TFA in water. The sample was eluted from the column with 0-80% (v/v) ethanol gradient (2 liters of 0, 20, and 35% ; 1.5 liters of 50%; and 1 liter of 80% ethanol). The 50% ethanol fraction having the most antimicrobial activity was concentrated in *vacuo* to about 20 ml at 30 °C. It was then mixed with an equivalent volume of acetonitrile (MeCN), and loaded onto a 5 X 25 cm Sephadex LH-60 (Sigma) column pre-equilibrated with 50% MeCN and 0.1% TFA in water. After elution, the bioactive fractions were pooled and concentrated to 10 ml. The final purification was accomplished by reverse-phase HPLC with a C8 VYDAC column (10 x 250 mm, 10-µm particle size, 300 Å pore size) using a gradient from 20 to 31% MeCN in 0.1% TFA and a flow rate of 2.5 ml/min.: Mass spectrometry of fresh CbnB2 was done by direct solution injection (50% aqueous MeCN, 0.1% TFA) using a VG Quattro triple quadruple instrument with an electrospray ionization source (Fisons, Manchester, United Kingdom). The identity of the purified CbnB2 was confirmed by mass spectrometry and co-injection on HPLC with an authentic sample of CbnB2. The molecular weight was found to be 4968.00 +/- 0.49 (cal. 4969.51, average of monoisotopes for the reduced form) and varied about one mass unit from run to run. Co-injection on HPLC gave a single peak with the same retention time as an authentic sample.

Reduction and Carboamidomethylation of CbnB2. Two identical samples were prepared by dissolving 50 µg of CbnB2 into 0.5 ml, pH 8.6, 0.6M tris-HCl buffer containing 6 M guanidinum hydrochloride. An aliquot of 20 µl of freshly prepared 280 mM DTT was added to one of the samples. After a brief period of vortexing, the sample was incubated at 37 °C for thirty minutes and then 150 µl of freshly prepared iodoacetamide (0.6 mM) was added dropwise over two minutes. The sample was then placed under an argon environment, in the absence of light, for one hour at 37 °C. The other sample was reacted with iodoacetamide under exactly the same conditions but without prior DTT reduction. The reaction mixture was directly loaded onto a reverse-phase HPLC (C8 VYDAC column, 4.6 x 250 mm, 5-µm particle size, 300 Å pore size, flow rate 1 ml/min) for further purification. After concentration *in vacuo,* the purified product was redissolved in aqueous MeCN (50 %) for mass spectrometry. The same experiments were also carried out on a synthetic version of the N-terminus of CbnB2 (residues 1-22). The peptide, in which both cysteines were reduced, was dissolved in a 0.1 M ammonium bicarbonate buffer (pH 8.0) and stirred under an oxygen environment overnight. The oxidized product was purified by reverse-phase HPLC as above. Two identical samples of both oxidized and reduced forms of the 22-mer were reacted with iodoacetamide as described above, one being treated with DTT while the other was not. Mass spectra of those samples were collected under the same conditions as for the full-length CbnB2 samples.

NMR Sample Preparation. A total of 5.0 mg of CbnB2 was dissolved in 700 μl of TFE-d_3/H_2O, 90%:10% v/v. The material was then sonicated (under dry N_2) for one minute using a Cole-Palmer sonicator to accelerate solvation. The final concentration of the peptide was 0.9 mM. The TFE-d_3 used in the sample preparation was 99.94% deuterated (Cambridge Isotopes, Woburn MA). All samples were acidified to pH 2.8 (uncorrected meter reading) by adding TFA (0.1% final concentration).

NMR Spectroscopy. All ^{1}H NMR experiments were performed on a Unity 500 or 600 MHz NMR spectrometer equipped with a 5-mm triple resonance probe. TOCSY experiments were carried out using the basic pulse sequence proposed by Bax and Davis (1985). Acquisition times were set to 0.2 seconds, relaxation delays were 2.0 seconds and spin-lock (MLEV-17) mixing times were 50 ms. NOESY (Jeener et al., 1979; Kumar et al., 1980) data were collected essentially identically to the TOCSY data, with mixing times ranging from 150 to 300 ms. DQF-COSY spectra were also collected using standard protocols (Piantini et al., 1982; Shaka and Freeman, 1983). All two-dimensional spectra were collected using 256 t_1 increments and spectral widths of 6000 Hz in both dimensions. To distinguish between overlapping signals, as well as to determine the relative amide ^{1}H exchange rates, additional TOCSY and NOESY spectra were collected at 10, 20, 25 and 30 ^{o}C (± 0.1 ^{o}C) respectively. Water suppression was achieved using solvent presaturation during the relaxation delay period. Prior to Fourier transformation, the data matrix was zero-filled to 4K x 4K complex points and multiplied by an approximate 90^o-shifted sine-bell squared weighting function in both dimensions. The residual signal from protonated TFE was used as a secondary chemical shift reference (3.88 ppm downfield from 2,2-dimethyl-2-silapentane-5-sulfonic acid (DSS) (Wishart et al., 1995)).

Complete ^{1}H NMR assignments for CbnB2 were obtained using well-established procedures (Wuthrich, 1986; Clore and Gronenborn, 1987). Individual spin systems were initially identified from TOCSY and COSY spectra and then sequentially assigned using NOESY spectra. $^3J_{HNH\alpha}$ coupling constants were obtained using line-width measurements from both TOCSY and NOESY spectra (Wang et al., 1997). Amide proton temperature coefficients were determined by measuring amide chemical shift changes from TOCSY and NOESY spectra collected over a range of temperatures (10 - 35 °C).

Structure Calculations for CbnB2. Interproton distance restraints were derived by analyzing previously assigned NOE peaks present in NOESY spectra acquired with 150 - 200 ms mixing times. Assigned NOE intensities, measured by volume integration, were classified into four groups (strong, medium, weak and very weak) corresponding to interproton distance restraints of 1.8-2.8 Å, 1.8-4.0 Å, 1.8-5.0 Å and 1.8-6.0 Å respectively. $^3J_{HNH\alpha}$ coupling constants measured from TOCSY and NOESY spectra using line-width measurements (Wang et al, 1997) yielded a total of 15 ϕ backbone torsion angles restraints. Backbone ϕ angle restraints were set to $-120 \pm 30°$ for those residues exhibiting large coupling constants ($^3J_{HNH\alpha} > 8.5$ Hz). For those residues in the well-defined central helical region (i.e. $^3J_{HNH\alpha} < 6.0$ Hz), their ϕ torsion angle was calculated using the Karplus equation (Vuister and Bax, 1993) and assigned a variance of ± 20°. Backbone ψ angle restraints were obtained from an analysis of the $d_{N\alpha}/d_{\alpha N}$ ratios (Gagne et al., 1994). For $d_{N\alpha}/d_{\alpha N}$ ratios less than one, the ψ angle restraint was set to $120 \pm 100°$. For $d_{N\alpha}/d_{\alpha N}$ ratios greater than one, the ψ angle restraint was set to $-30 \pm 110°$. During the calculation, all ω angles were set to $180 \pm 10°$. Backbone hydrogen bonds were identified by analyzing the NH temperature coefficients and comparing them to the previously derived secondary structure. Each hydrogen bond was defined using two distance restraints, $d_{O\text{-}H}$=1.8-2.4 Å and $d_{O\text{-}N}$=2.7-3.5Å. The disulfide bond between Cys9 and Cys14 was defined by three distance restraints, $d_{S(i)\text{-}S(j)} = 2.02 \pm 0.05$ Å, $d_{S(i)\text{-}C\gamma(j)} = 2.99 \pm 0.05$ Å, and $d_{C\gamma(i)\text{-}S(j)} = 2.99 \pm 0.05$ Å. Force constants for NOE-derived distance restraints were set to 50 kcal mol^{-1} Å^{-2}. Dihedral angle force constants were initially set to 5 kcal mol^{-1} rad^{-2} during the high-temperature dynamics run and increased to 200 kcal mol^{-1} rad^{-2} during the annealing phase. In total, 319 NOE distance restraints (145 intra-residue, 89 sequential, 63 medium range and 22 hydrogen bond-derived restraints), 88 dihedral angle restraints (15 ϕ angle restraints, 22 ψ angle restraints and 47 ω angle restraints), 8 $^3J_{HNH\alpha}$ coupling constant restraints and 163 proton chemical shift restraints were used to generate our ensemble of CbnB2 structures. This represents an average of 13 restraints/residue.

During the first set of structure calculations, only NOE derived interproton distance restraints were included as experimental input. Thirty structures were generated using a simulated annealing protocol (Nilges et al., 1988) with 12000 high-temperature steps (60 ps at 1000 K) followed by 6000 cooling steps (30 ps, final temperature of 100K), as implemented in X-PLOR (Version 3.8.5 - Brunger, 1992). Further refinement was carried out using the same protocol but with 6000 high-temperature steps (30 ps) and 4000 cooling steps (20 ps). Those structures having no in inter-proton distance restraint violations greater than 0.5 Å were accepted and used as input structures for the second (torsion angle) phase of refinement. For the final stage of structural refinement, those structures exhibiting no interproton distance and torsion angle restraint violations greater than 0.5 Å and 5° respectively, were accepted and further refined against the measured $^3J_{HNN\alpha}$ coupling constants (Garrett et al., 1994) and proton chemical shift restraints (Kuszewski et al., 1995). During this final refinement stage, 800 steps of conjugate-gradient minimization were performed. The force constants for $^3J_{HNN\alpha}$ coupling constant restraints and proton chemical shift restraints were 1.0 kcalmol^{-1}Hz^{-2} and 7.5 kcalmol^{-1}ppm^{-2}.

Structure Refinement of LeuA. To complete the structural refinement of LeuA, previously collected TOCSY and NOESY spectra (Fregeau et al, 1997) were reanalyzed. Based on the previously published chemical shift assignments for LeuA, $^3J_{HNH\alpha}$ coupling constants were measured (via linewidth measurements) and additional NOE peaks were assigned from the NOESY spectrum. The 18 structures previously generated by Fregeau et al. (Fregeau et al., 1997; PDB accession #: 2LEU) were used as initial input structures. These structures were first refined against the newly assigned interproton distance restraints using the previously described simulated annealing protocol and then further refined against $^3J_{HNN\alpha}$ coupling constants and proton chemical shifts. A total of 434 interproton distance restraints (155 intra-residue, 118 sequential, 153 medium and long-range NOE derived distance restraints as well as 8 hydrogen bond derived distance restraints), 27 $^3J_{HNN\alpha}$ coupling constant restraints and 136 proton chemical shift restraints (representing an average of 16 restraints/residue) were used during the refinement of LeuA. This compares to an average of just 11 restraints/residue (407 distance restraints) which were used in generating the initial LeuA structure (Fregeau et al., 1997).

Structure Evaluation and Analysis. The quality of the final ensemble of both sets of CbnB2 and LeuA structures was assessed using the program PROCHECK-NMR version 3.4.4 (Laskowski et al., 1996). MOLMOL (Koradi et al., 1996) was used to visualize, superimpose and calculate RMSD values for all structures. Homology modeling of CbnB2 was conducted using the WHATIF on-line sever (Vriend, 1990; http://swift.embl-heidelberg.de). Theoretical ^{1}H chemical shifts were calculated using the program TOTAL (Williamson et al., 1995). Dihedral angles were extracted using the program VADAR (Wishart et al., 1995) and sequence dependent secondary structure analysis was performed using PepTool (Wishart et al., 1997).

5.3. Results and Discussion

Redox Statues of CbnB2. Earlier reports (Quadri et al., 1997) had indicated that CbnB2 was apparently unique among Type IIa bacteriocins in that it did not posses an N-terminal disulfide bond. The ambiguous NMR results concerning the structure and oxidation state of this region led us to reinvestigate the redox state of CbnB2's two cysteines using other means, including kinetic studies of the redox reaction of CbnB2 and LeuA peptide mimics, mass spectroscopy, and thiol modification experiments.

Based on recent studies of *E. coli* thioredoxin where an unusual Lys-Cys interaction has been reported (Dyson et al., 1997), we hypothesized that Lys13 in CbnB2 may undergo a similar lysine-thiol interaction, which could prevent facile oxidation of the disulfide bridge between the Cys9 and Cys14. To test this hypothesis, a model system for mimicking these two bacteriocins was prepared by synthesizing six different hexapeptides corresponding to the region bounded by the two cysteines in CbnB2 and Leu A. Two of these peptides (peptides #1 and #2, see Table 5.1) were identical in sequence to residues 9 - 14 in native LeuA and CbnB2. The other four were homologous to the CbnB2 sequence. Kinetic studies of the reduction of these cyclic peptides in the presence of excess DTT showed second-order reaction kinetics at pH 5.0. As expected, reoxidization with 50 fold excess oxygen obeys first order kinetics. The measured reduction and oxidation rate constants for all six peptides are listed in Table 5.1. This data indicates that there is no obvious evidence for Lys13 either decreasing the oxidization rate or increasing the reduction rate. Therefore, our measured redox reaction constants for these model peptides does not indicate a significant Lys-thiol interaction.

Reduction and Carboamidomethylation of CbnB2. As outlined in Figure 5.1, the redox status of CbnB2 was further investigated using carboamidomethylation of the CbnB2 cysteines. The molecular weights of the CbnB2 carboamidomethylation products clearly show that CbnB2 could not be carboamidomethylated if the sample was not first reduced with DTT. Among those samples which could not be carbonamidonmethylated, the oxidized form of CbnB2 was recovered each time. Because carboamidomethylation only occurred when the CbnB2 sample was reduced, this clearly indicates that a disulfide bond existed between the two cysteines of CbnB2 (Cys9 and Cys14) prior to reduction. Furthermore, activity studies show that the recovered unmodified CbnB2 always maintains its antimicrobial activity, while the modified product does not. For the oxidized synthetic 22-mer (corresponding to the N terminus of CbnB2), the same modification patterns were observed. Specifically, the two cysteine thiol groups were carboamidomethylated only if the sample was first reduced with DTT. For the reduced form of the 22-mer, the two cysteines could be carboamidomethylated regardless of whether or not DTT was used to effect disulfide reduction.

Chemical Shift Assignment and Secondary Structure of CbnB2. A complete list of the assignments, including the $^3J_{HNN\alpha}$ coupling constants and amide proton temperature coefficients for CbnB2 is given in Table 5.2. The observed medium-range (i, i+2; i, i+3) NOEs and strong amide proton (i, i+1) NOEs and characteristically small $^3J_{HNN\alpha}$ coupling constants from Trp18 to Ser39 indicate that CbnB2 adopts a helical configuration in this region. A section of the NOESY spectrum, collected at 25 °C, corresponding to this helical segment is shown in Figure 5.2.

It is well known that ^{1}H NMR chemical shifts are strongly dependent on the character and nature of protein secondary structure (Wishart et al., 1991). In particular, it has been found that α-^{1}H chemical shifts experience an upfield shift (with respect to the random coil value) when in a helical configuration and a comparable downfield shift when in a β-strand or extended configuration (Wishart et al., 1991, 1992 and 1995; Williamson, 1990). Net structural chemical shifts (the observed chemical shift minus the random-coil value - Wishart and Nip, 1998) for CbnB2's α-protons are shown in Figure 5.3b. The nearly continuous set of upfield α-proton chemical shifts from Trp18 to Ser39 clearly confirm the existence of an α helix in this region. Small amide temperature coefficients (<3 ppb/°C), also observed over the same region, indicate the presence of extensive hydrogen bonding and are further suggestive of a well-defined α-helical structure.

Figure 5.3b also shows that the α-protons for several N-terminal residues (8-17) exhibit a slight downfield shift but one that is not significant enough to indicate any well-defined β-sheet structure. Several weak NOEs between the Val7 γCH_3 protons and the Tyr3 ring protons were also observed, but no other significant medium or long-range NOEs could be found in the N-terminus. These data, along with the random coil-like $^3J_{HNN\alpha}$ coupling constants (~ 7 Hz) measured throughout this region, indicate that the N-terminus of CbnB2 is essentially unstructured. It is also worth noting that no NOEs (at any temperature) could be detected between Cys 9 and Cys 14 or any other residues involved in the N-terminal disulfide bond. This paucity of NOE data made it essentially impossible to determine if CbnB2 was, in fact, reduced or oxidized by NMR. Confirmation of the oxidation state of our NMR sample by subsequent mass spectrometry and chemical modification experiments (*vide infra*) eventually allowed a disulfide constraint to be built into our final model. Evidently the rapid motions experienced by the N-terminus must have led to a time-averaged dissipation of the expected NOEs (Ikura et al., 1991). As for CbnB2's C-terminus (residues 41-48), the absence of any medium or long range NOEs, the "random-coil" $^3J_{HNN\alpha}$ coupling constants (~ 7 Hz) and featureless net structural α-proton chemical shifts, indicate that this region, too, is largely unstructured. The lack of structure for both the N and C terminus was confirmed by additional TOCSY and NOESY measurements collected over a range of temperatures (10 °C - 35 °C). These latter results indicate that the disordered structure observed for the N terminus of CbnB2 at 25 °C is not necessarily a function of the spectral collection conditions, but rather a function of the inherent structural propensity of this region.

Three-dimensional Structure of CbnB2. A summary of the structural statistics for the final set CbnB2 structures is provided in Tables 5.3 and 5.4. The criteria for acceptance for this final set of structures were as follows: no interproton distance restraint violation greater than 0.5 Å, no dihedral angle restraint violation greater than 5°, no RMS deviation from ideal bond lengths greater than 0.02 Å, no RMS deviation from ideal bond angles greater than 2.0°, no coupling constant restraint violation greater than 1 Hz, and no more than 14 proton chemical shift restraint violations (i.e. $|\delta_{obs}-\delta_{cal}| > 0.5$ ppm) for any given structure. For some structures, several rounds of refinement were required to meet the above criteria. Twenty of the initial 30 structures eventually satisfied these criteria and were accepted into the final ensemble of CbnB2 solution structures. For the helical region (residues 19 to 39), the average root mean square deviation (RMSD) relative to the mean was calculated to be 0.48 Å for the backbone atoms, and 0.94 Å for all heavy atoms. Closer examination of this helix indicates that the side chains of several hydrophobic residues Trp18, Phe22, Tyr26, Ile30, Phe33, Val34 and Val37 form a long, well-defined hydrophobic surface on one side of the helix. Conversely, the side-chains of several hydrophilic residues Asn17, Gln20, Glu24, Arg25, Asn31, Ser35, and Ser39 form a well-defined hydrophilic surface on the opposite side.

The amphipathic character of many antimicrobial peptides is thought to play an important role in their antimicrobial activity (Nissen-Meyer and Nes, 1997; Fimland et al., 1998) and CbnB2 is no exception. Earlier studies have already shown that the replacement of just one hydrophobic residue (Phe33) with a more hydrophilic residue (Ser) completely inactivates CbnB2 (Quadri et al., 1997). In addition to the importance of amphipathicity, the presence of positively charged amino acids in antimicrobial peptides is also believed to be critical to their cell targeting and cell specificity. Cationic residues are thought to be involved in the initial interaction with anionic phospholipids, which are present in relatively high concentrations in many bacterial membranes (Jacob and Zasloff, 1994). With five positively charged amino acids (Lys11, Lys13, Arg25, Arg46 and Arg47) and just one negatively charged amino acid (Glu24), CbnB2 exhibits the characteristically strong net positive charge of many antimicrobial peptides. However, it is notable that only Arg25 lies in the amphipathic helical region. Meanwhile, Lys11 and Lys13, which reside on the

N terminus, and Arg46 and Arg47, which reside on the C terminus, could potentially facilitate key interactions with anionic bacterial membranes.

As discussed previously, Type IIa bacteriocins are characterized by the presence of a highly conserved N-terminal sequence motif: YGNGVXC (Klaenhammer, 1993, Bennik et al., 1998). This conserved N-terminal region has been suggested to form part of a recognition sequence for a putative membrane-bound protein receptor (Fimland et al., 1996; Fleury, 1996). Consequently it was of some interest to study the three-dimensional structure of this region in detail to see if there is a correspondingly well conserved structural motif. As mentioned earlier, the NOEs observed between the Val7 γCH_3 protons and the Tyr3 ring protons indicate the existence of a weak, or poorly defined reverse turn in this region. While the overall structure of the N-terminus is highly disordered, we find that when we superimpose just this short motif region (residues 2 to 7), the average RMSD relative to the mean coordinates is 0.99 Å for the backbone atoms, and 1.91 Å for heavy atoms. This modest fit indicates that at least some nascent turn structure is present in this putative receptor binding region. Interestingly, the proximity of the Tyr3 and Val7 side chains in this turn leads to the formation of a small hydrophobic surface which may have some impact on the peptide's ability to interact with the cell membrane. In particular, it has recently been shown that the substitution of Tyr3 with Phe can reduce the antimicrobial activity of CbnB2 by 7 fold (Quadri et al., 1997). A steroview of the backbone superimposition of the N-terminus of CbnB2 and LeuA is shown in Figure 5.4.

Refined Three-dimensional Structure of LeuA. Because of the importance of conducting a detailed comparison between the LeuA and CbnB2 solution structures, we decided to refine the previously published LeuA structure (Fregeau et al., 1997) by taking advantage of recent advances in NMR structure refinement protocols (Garrett et al., 1994; Kuszewski et al., 1995) and coupling constant measurements (Wang et al., 1997). During the course of this LeuA refinement, 190 new restraints were added, including 30 medium and long range NOE-derived distance constraints, 27 $^3J_{HNN\alpha}$ coupling constant restraints and 136 proton chemical shift restraints. As can be seen in Figure 5.5, the addition of these restraints substantially improved the LeuA structure in almost every respect. Relative to the earlier set of LeuA structures, the RMS deviation of the backbone and heavy atoms fell from 1.10 Å and 1.57 Å (for the unrefined set) to just 0.67 Å and 1.04 Å (for the refined set). Using standard PROCHECK structure "quality" statistics, the overall G-factor increased from -0.7 (for the unrefined set) to just -0.1 (for the refined set). Furthermore, for the newly refined set of LeuA structures, 75% of the ϕ/ψ dihedral angles were located within the "most favorable" region of Ramachandran space while just 65% fell into this region for the unrefined set. The NMR structural statistics for LeuA are provided in Table 5.3. Equally important, the calculated chemical shifts for the N-terminal region of the refined set of LeuA structures now shows a substantially better agreement to the observed chemical shifts than for the unrefined set (The RMSD between the observed and predicted α-proton chemical shifts is 0.27 ppm for the refined set versus 0.42 ppm for the unrefined set).

Despite the evident improvement in the precision and accuracy of the LeuA structure, it is clear that the overall fold of LeuA is essentially the same as previously described (Fregeau et al., 1997). In particular, LeuA still adopts a well-defined tertiary structure from residues 2 to 31 including a 3-stranded antiparallel β-sheet (residues 2-16) followed by an amphiphilic helix (residues 17-31). The β-sheet and α–helix conformation is clearly characterized by the net structural α-proton chemical shifts shown in Figure 5.3a.

Comparison between CbnB2 and LeuA. As our electrospray MS and thiol modification experiments has shown, both CbnB2 and LeuA share a disulfide bond between their conserved cysteines (Cys9 and Cys14). Furthermore, both peptides share a well-defined amphipathic α-helix located near their C termini (residues 19-39 for CbnB2; residues 17-31 for LeuA). These results are consistent with their high level of sequence homology and their presumably common mode of action. However, what is quite unexpected is fact that the N-terminal regions of CbnB2 and LeuA exhibit significant differences both in calculated tertiary structure and measured chemical shift. A superimposed ribbon diagram of CbnB2 and LeuA is shown in Figure 5.6. From this diagram it is clear that the CbnB2 N-terminus is highly disordered while the corresponding region in LeuA forms a well defined antiparallel β-sheet. The structural difference between these two peptides is further demonstrated by plotting their α-proton chemical shift differences (Figure 3.3c).

What is particularly striking is that these structural differences persist despite a very high level of shared sequence similarity between the two peptides (66% sequence identity over the first 24 residues).

One of the central dogmas of structural biology is that shared sequence homology almost always implies shared structural homology (Doolittle and Feng, 1987). Of course the actual level and extent of sequence homology is critical to determining the level and extent of observed structural homology (Kabsch and Sander, 1985; Chothia and Lesk, 1986). Typically, short regions of sequence homology require higher levels of sequence identity to imply or confer structural conservation (Chothia and Lesk , 1986). There are many instances, for example, where identical penta- and hexapeptides have been found to have completely different tertiary structures (Kabch and Sander, 1985; Wodak and Rooman, 1990). There are also several documented cases where single amino acid substitutions over longer stretches of amino acids have led to significantly different structures. For example, bovine and pancreatic phospholipase share a segment of 13 amino acids that differs by just one residue (LDSCKV(F)LVDNPYT) but X-ray analysis of these two closely related enzymes indicates this stretch of amino acids assumes two quite different conformations (Kabsch and Sander, 1984 and 1985). Nevertheless, empirical studies by Sander and Schneider (1991) indicate that for an alignment spanning 24 residues, a level of sequence identity of 48% should be enough to confer structural homology. In the case of CbnB2 and LeuA, the level of sequence identity for the first 24 residues is actually 66%.

To understand the possible causes for these structural differences in the N-termini of CbnB2 and LeuA, we believed it was critical to rule out any sources of experimental error that may have led to the generation of incorrect or unreasonable structures. In particular we investigated three possibilities: (1) differential thermal stability between LeuA and CbnB2; (2) errors in the assignments or structure of LeuA; and (3) errors in the assignments or structure of CbnB2. Differential thermal stability investigated by carefully checking the NMR data collected for CbnB2 at two lower temperatures (10 °C and 15 °C) and looking for evidence of additional or previously undetected structure. No significant differences in chemical shifts, coupling constants or NOE patterns were found. This suggests that the disorder seen in the N terminus of CbnB2 is not a consequence of thermal denaturation or low thermal stability. Similarly, we investigated whether the 3-stranded β-sheet in LeuA was particularly temperature sensitive (i.e. thermally unstable). Chemical shift and NOE measurements at 25 °C and 35 °C indicated that this structure in LeuA is, in fact, quite stable. These results suggest that differences in thermal stability is not the primary reason for the observed structural differences in LeuA and CbnB2.

Having found that these differences could not be attributed to differential thermal stability, we subsequently investigated whether there may be errors in either the published assignments or published structure for LeuA (Fregeau et al., 1997). While we did detect some errors, these were corrected during our subsequent structural refinement and, as stated earlier, the refined structure of LeuA closely resembles the previously published version - including the well-defined N-terminal three-stranded β-sheet. In light of this result, we decided to investigate the last possible source of error: incomplete structural sampling during the generation of the CbnB2 structures.

When generating the solution structure of a relatively flexible region of a peptide (such as the N-terminus of CbnB2), it is always possible that the structure generation protocol does not completely sample all parts of conformation space. This is called incomplete sampling. Consequently, it may be possible that a structure in the N terminus of CbnB2, identical to the one in LeuA, is completely compatible with the observed NMR data. To investigate this possibility we decided to "thread" (Bryant and Lawrence, 1993) the CbnB2 sequence into the refined LeuA structure and compare the calculated NMR data for this LeuA-like structure with the observed NMR data Specifically, model structures of the first 37 residues of CbnB2 were built (via WHATIF; Vriend, 1990) using our previously refined LeuA structure as a template. The correctness of this "remodeled" CbnB2 was then evaluated by comparing its predicted ^{1}H chemical shifts with the observed values using the program TOTAL (Williamson et al., 1995). Recently developed chemical shift calculation methods (such as TOTAL) have made it possible to accurately (±0.25 ppm) calculate ^{1}H chemical shifts for proteins having well resolved structures (Osapay and Case, 1991; Williamson et al., 1995). The calculated α-proton chemical shifts together with the observed values for these "remodeled" CbnB2 structures is plotted in Figure 5.7. From these two figures it is clear that unrealistically large α-proton chemical shift differences (greater than two standard deviations) are predicted for residues Val7, Ser8, Cys9, Cys14, and Val16 of the "remodeled" CbnB2 structure. Likewise, significant $^3J_{HNN\alpha}$ coupling constant differences are also predicted for residues Asn5, Lys13 in the same remodeled structure. These results indicate that a three-stranded β-sheet (such as the one found in LeuA) is structurally incompatible with the observed NMR data for CbnB2. In other words, we can rule out incomplete conformational sampling as a possible reason for the structural differences observed between LeuA and CbnB2.

Why are the N-termini Different? While there are many examples in the literature where single amino acid substitutions in peptides and proteins have led to profound changes in function (Barden et al., 1997; Kippen et al., 1994; Demene et al., 1994; Volkman and Wemmer, 1997; Yi et al., 1994; Young et al., 1998), these substitutions usually lead to complete denaturation or complete loss of activity. What is particularly interesting about the situation with LeuA and CbnB2 is that the observed structural differences do not seem to affect their strong antimicrobial activity. With essentially all reasonable sources of experimental error excluded, we believe that the best explanation for these profound structural differences must lie in the intrinsic nature of the two peptides and/or their surrounding solvent (TFE).

With respect to the intrinsic nature of these peptides, analysis of the secondary structural propensity of the N-terminus of LeuA and CbnB2 reveals three things. First, the average β-sheet propensity, P_β (Kim and Berg, 1993), from residues 2 to 16 is slightly higher than for LeuA than CbnB2 (1.05 vs. 1.00). Secondly, standard secondary structure prediction methods (Chou and Fasman, 1974; 1978; Garnier et al., 1978; 1996; Levin and Garnier, 1988) obtained from PepTool (Wishart et al., 1997) and the PHD secondary structure server (Rost and Sander, 1993; 1994) predict that LeuA is somewhat more likely to form the central β-strand (7-10) than CbnB2 (data not shown). And third, it has been shown that β-turn propensity can have a strong effect on the stability of both β-hairpins and β-sheets (Blanco et al., 1994, Wishart et al., 1995; Najbar et al., 1997). In this regard, the differing amino acids at positions 12 and 13 (Ser12/Gly13 for LeuA; Thr12/Lys13 for CbnB2), corresponding to the i+2 and i+3 positions of the LeuA β-turn may have an important effect. More specifically, the β-turn probability (Chou and Fasman, 1978) for this particular turn is calculated to be 1.88 x 10^{-4} in LeuA while it is just 0.85 x 10^{-4} for CbnB2.

While linear sequence analysis can reveal some interesting differences, a more telling picture can be gained by analyzing the tertiary structure of LeuA. A close look at the β-sheet formed at the N-terminus of Leu A reveals that it is quite amphipathic. Indeed, four residues: Tyr2, His8, Thr10 and Ala 24 form a hydrophobic cluster on one side of the β-sheet, while four residues: Lys1, Asn5, Cys9 and Cys14 form a weakly hydrophilic cluster on the other side (Figure 5.9). However, such an amphipathic β-sheet structure could not form in CbnB2 if it adopted the same fold as LeuA. This is because all four hydrophobic residues (Tyr2, His8, Thr10 and Ala24) in LeuA have been replaced by hydrophilic or neutral residues (Asn2, Ser8, Ser10, and Glu24) in CbnB2. The significant hydrophobic/hydrophilic differences arising from these four key residue substitutions could account, at least in part, for the observed structural difference between LeuA and CbnB2.

The amphipathic nature of the LeuA β-sheet may also play into the hand of how TFE maintains or induces structure in this peptide. Solvent effects have long been known to play a very important role in determining the secondary structure of an amino acid sequence. Indeed, solvent effects can often override the intrinsic secondary structural propensity for almost any given sequence (Zhong and Johnson, 1992; Waterhous et al., 1994; Najbar et al., 1997). It has been proposed that the structure stabilizing properties of TFE arise from two important physical properties, namely the hydrophobicity of the trifluoromethyl group and the hydrophibicity (or hydrogen bonding character) of the hydroxyl group (Rajan and Balaram, 1996). In particular, TFE provides a more hydrophobic and less basic environment than pure water. The decreased capacity of TFE to accept hydrogen bonds is thought to weaken the peptide-solvent interactions, leading to stronger intra-molecular interactions and increased formation of secondary structure (Storrs et al., 1996; MacPhee et al., 1997; Luo and Baldwin, 1997). Because of its intrinsic amphipathic or bipolar character, TFE is particularly effective at inducing or stabilizing amphipathic structures, especially α-helices (Nelson and Kallenbach, 1989; Merutka and Stellwagen, 1989; Sonnichsen et al., 1992, Zhou et al., 1993). At the same time, amphipathic β-sheets and β-hairpins have also been shown to be stabilized by TFE (Mutter et al., 1985; Blanco et al., 1994; Wang et al., 1995; Wishart et al., 1995; Ramirez-Alvarado et al., 1996; Fregeau et al., 1997).

On the other hand, if an amino acid sequence has very little potential to form an amphipathic structure or it has very little intrinsic secondary structure propensity, TFE will not typically induce a well-defined structure (Sonnichsen et al., 1992; Hamada et al., 1995; Arunkumar et al., 1996; Rajan and Balaram, 1996; Luidens et al., 1996; Luidens et al., 1996; Dong et al., 1998). These results suggest that the effects of TFE on β-sheet stability is both sequence and structure dependent, with the structure dependence being based on a sequence's ability to form an amphipathic surface. In this regard, the amphipathic structure formed at the N-terminus of LeuA could be a major reason why the observed β-sheet is stabilized in TFE. This also suggests that the LeuA β-sheet, which is only seen in TFE, may in fact be an artifact of the solution (i.e. TFE) conditions. This further implies that the N-terminal structure (i.e. the presumptive receptor binding site) of type IIa bacteriocins may not be central to their function. Detailed structural studies of other type IIa bacteriocins will likely be needed to resolve this issue.

5.4. Conclusion

Our results clearly show that, like all other type IIa bacteriocins, CbnB2 possesses a disulfide bond between Cys9 and Cys14. Furthermore, CbnB2 and LeuA, despite sharing considerable sequence similarity (66.6% for the first 24 residues), exhibit very different N-terminal structures. LeuA adopts a well-defined 3-stranded antiparallel β-sheet; while CbnB2 adopts a disordered structure. This structural difference between CbnB2 and LeuA can be attributed to a combination of differing secondary structural propensities and the differing effects of TFE on amphipathic structures. This study also indicates that while the N-terminal sequence may be important, the β-sheet found at the N-terminus of LeuA may not be critical to the antimicrobial functions of type IIa bacteriocins.

References

Arunkumar, A. I., Kumar, T. K., Jayaraman, G. and Samuel, D. and Yu, C (1996) *J. Biomol.Struct. Dyna.* 14, 381-385.

Avelj, F. and Moult, J. (1995) *Biochemistry* 34, 755-764.

Avbelj, F. and Fele, L. (1998) J. *Mol. Biol.* 279, 665-684.

Bai, Y. and Englander, S. W. (1994) *Proteins: Struct. Funct. Genet.* 18, 262-266.

Barden, J. A., Cuthbertson, R. M., Wu, J. Z., Moseley, J. M. and Kemp, B. E. (1997) *J. Biol. Chem.* 272, 29572-29578.

Bax, A. and Davis, D.G. (1985) *J. Magn. Reson.* 65, 355-360.

Bennik, M. H. J., Vanloo, B., Brasseur, R., Gorris, G. M. and Smid, E. J. (1998) *Biochim. Biophys. Acta* 1373, 7-58.

Blaber, M., Zhang, X., Lindstrom, J. D., Pepiot, S. D., Basse, W. A. and Matthews, B. W. (1994) *J. Mol. Biol.* 235, 600-624.

Blanco, F. J., Jimenez, M. A., Pineda, A., Rico, M., Santor, J. and Nieto, J. L. (1994) *Biochemistry* 33, 6004-6014.

Boman, H. G. (1995) *Annu. Rev. Immunol.* 13, 61-92.

Brunger, A. T. (1992) X-PLOR Version 3.1. *A System for X-ray Crystallography and NMR,* Yale University Press, New Haven, CT.

Cammue, B.P.A., De Bolle, M.F. C., Schoffs, H.M.E., Terras, F.R.G., The-vissen, K., Osborn, R.W., Rees, S.B. and Broekaert, W.F. (1994) In: Bomam, H.G., Marsh, J. and Goode, J.A. (Eds) *Antimicrobial peptides.* Wiley, New York, pp 91-106.

Chothia, C. and Lesk, A. M. (1986) *EMBO J.* 5, 823-826.

Chou, P. Y. and Fasman, G. D. (1974) *Biochemistry* 13, 211-222.

Chou, P. Y. and Fasman, G. D. (1977) *J. Mol. Biol.* 115, 135-175.

Chou, P. Y. and Fasman, G. D. (1978) *Annu. Rev. Biochem.* 47, 251-276.

Clore, G. M. and Gronenborn, A. M. (1987) *Prot. Eng.* 1, 275-288.

Creamer, T. P. and Rose, G. D. (1995) *Prot. Sci.* 4, 1305-1314.

Davies, J. (1994) *Science* 264, 375-381.

Demene, H., Dong, C. Z., Ottmann, M., Rouyez, M. C., Jullian, N., Morellet, N., Mely, Y., Darlix, J. L., Fournie-Zaluski, M. C. and Saragosti, S. et al. (1994) *Biochemistry* 33, 11707-11716.

Delves-Broughton, J., Blackburn, P., Evans, R. J. and Hugenholtz, J. (1996) *Antonie Van Leeuwenhoek* 69, 193-202.

Doolittle, R. F. and Feng, D. F. (1987) *Cold Spring Harbor Symposia on Quantitative Biology* 52, 869-874.

Dong, A., Matsuura, J., Manning, M. C. and Carpenter, J. F. (1998) *Archives of Biochemistry and Biophysics* 355, 275-281.

Erickson, B. W. and Merrifield, R. B. (1976) in *The Proteins* (Neurath, H. and Hill, R. H., eds) Vol. II, pp 255-527, Academic Press, New York.

Esler, W. P., Stimsion, E. R., Ghilardi, J. R., Lu, Yi-An, Felix, A. M., Vinters, H. V., Mantyh, P. W., Lee, J. P. and Maggio, J. E. (1996) *Biochemistry* 35, 13914-13921.

Fimland, G., Blingsmo, O. R., Sletten, K., Jung, G., Nes, I. F. and Nissenmeyer, J. (1996) *Appl. Environ. Microb.* 62, 3313-3318.

Fimland, G., Jack, R., Jung, G. , Nes, I. F. and Nissenmeyer, J. (1998) *Appl. Environ. Microb.* 64, 5057-5060.

Fleury, Y., Dayem, M. A., Montagne, J. J., Chaboissseau, E., Le Caer, J. P., Nicolas, P. and Delfour, A. (1996) *J. Biol. Chem.* 271, 14421-14429.

Fregeau, G. N. L., Sailer, M., Niemczura, W. P., Nakashima, T. T., Stiles, M. E. and Vederas, J. C. (1997) *Biochemistry* 36, 15062-15072.

Gagne, S. M., Tsuda, S., Li, M. X., Chandra, M., Smille, L. B. and Sykes, B. D. (1994) *Prot. Sci.* 3, 1961-1974.

Garrett, D. S., Kuszewski, J., Hancock, T. J., Lodi, P. J., Vuister, G. W., Gronenborn, A. M. and Clore, G. M. (1994) *J. Magn. Reson.,* Series B 104, 99-103.

Garnier, J., Osguthorpe, D. J. and Robson, B. (1978) *J. Mol. Biol.* 120, 97-120.

Garnier, J., Gibrat, J. F. and Robson, B. (1996) *Meth. Enzymol.* 266, 540-553.

Hamada, D., Kuroda, Y., Tanaka, T. and Goto, Y. (1995) *J. Mol. Biol.* 254, 737-746.

Hastings, J. W., Sailer, M., Johnson, K., Roy, K. L., Vederas, J. C. and Stiles, M. E. (1991) *J. Bacteriol.* 173, 7491-7500.

Bryant, S.H. and Lawrence, C.E. (1993) *Proteins: Struct. Func. Genet.* 16, 92-112.

Ikura, M., Kay, L. E., Krinks, M. and Bax, A. (1991) *Biochemistry* 30, 5498-5504.

Jacob, L. and Zasloff, M. (1994) In: Bomam, H.G., Marsh, J. and Goode, J.A. (eds) *Antimicrobial Peptides.* Wiley, New York, 91-106.

Jasanoff, A. and Fersht, A. R. (1994) *Biochemistry* 33, 2129-2135.

Jeener, J., Meier, B. H., Bachmann, P. and Ernst, R. R. (1979) *J. Chem. Phys.* 71, 4546-4553.

Kabsch, W. and Sander, C. (1984) *Proc. Natl. Acad. Sci. U.S.A* 81, 1075-1078.

Kabsch, W. and Sander, C. (1985) *Nature* 317, 207.

Kippen, A. D., Arcus, V. L. and Fersht, A. R. (1994) *Biochemistry* 33, 10013-10021.

Kim, C. A. and Berg, J. M. (1993) *Nature* 362, 267-270.

Klaenhammer, T. R. (1993) *FEMS Microb. Rev.* 12, 39-86.

Koradi, R., Billeter, M. and Wuthrich, K. (1996) *J. Mol. Graphics.* 14, 51-55.

Kumar, A. Ernst, R. R. and Wuthrich, K. (1980) *Biochem. Biophys. Res. Commun.* 95, 1-6.

Kuszewski, J., Gronenborn, A. M. and Clore, G. M. (1995) *J. Magn. Reson.*, Series B 107, 293-297.

Laskowski, R. A., Rullmannn, J. A., MacArthur, M. W., Kaptein, R. and Thornton, J. M. (1996) *J. Biomol. NMR.* 8, 477-86.

Levin, J. M. and Garnier, J. (1988) *Biochim. Biophys. Acta* 955, 283-95.

Luidens, M. K., Figge, J., Breese, K. and Vajda, S. (1996) *Biopolymers* 39, 367-376.

Luo, E. and Baldwin, R. L. (1997) *Biochemistry* 36, 8413-8421.

MacPhee, C. E., Perugini, M. A., Sawyer, W. H. and Howlett, G. J. (1997) *FEBS Lett.* 416, 265-268.

Meienhofer, J., Waki, M., Heimer, E. P., Lambros, R. J., Makofske, R. C. and Chang, C.D. (1979) *Int. J. Pept. Prot. Res.* 13, 35-42.

Merutka, G. and Stellwagen, E. (1989) *Biochemisty* 28, 352-357.

Mutter, M., Maser, F., Altmann, K. H., Toniolo, C. and Bonora, G. M. (1985) *Biopolymers* 24, 1057-1074.

Najbar, L. V., Craik, D. J., Wade, J. D., Salvatore, D. and McLeish, M. J. (1997) *Biochemistry* 36, 11525-11533.

Narayanan, U., Keiderling, T. A., Bonora, G. M. and Toniolo, C. (1986). *J. Am. Chem. Soc.* 108, 2431-2437.

Nilges, M., Clore, G. M. and Gronenborn, A. M. (1988) *FEBS Lett.* 229, 317-324.

Nikaido, H. (1994) *Science* 264, 362-388.

Nissen-Meyer, J. and Nes, I. F. (1997) *Arch. Mirobiol.* 167, 67-77.

Osapay, K. and Case, D. A. (1991) *J. Am. Chem. Soc.* 113, 9436-9444.

Osapay, K. and Case, D. A. (1991) *J. Biomol. NMR* 4, 215-230.

Piantini, U., Soresen, O.W. and Ernst, R. R. (1982) *J. Am. Chem. Soc.* 104, 6800-5.

Quadri, L. E. N., Sailer, M., Roy, K. L., Vederas, J. C. and Stiles, M. E. (1994) *J. Biol. Chem.* 269, 12204-12211.

Quadri, L. E. N., Sailer, M., Terebiznik, M. R., Roy, K. L., Vederas, J. C. and Stiles, M. E. (1995) *J. Bacteriol.* 177, 1144-1151.

Quadri, L. E. N., Yan, L. Z., Stiles, M. E. and Vederas, J. C. (1997) *J. Biol. Chem.* 272, 3384-3388.

Rajan, R. and Balaram, P. (1996) *Int. J. Pept. Prot. Res.* 48, 328-336.

Ramirez-Alvarado, M., Blanco, F. J. and Serrano, L. (1996) *Nature Struct. Biol.* 3, 604-614.

Richardson, J. S. (1977) *Nature* 268, 495-500.

Robinson, W. E. Jr., McDougall, B., Tran, D. and Selsted, M. E. (1998) *J. Leuk. Biol.* 63, 94-100.

Rooman, M. J., Rodrigues, J. and Wodak, S. J. (1990) *J. Mol. Biol.* 213, 337-350.

Rost, B. and Sander, C. (1993) *J. Mol. BIol.* 232, 584-599.

Rost, B. and Sander, C. (1994) *Proteins* 20, 584-599.

Sailer, M., Helms, G. L., Henkel, T., Niemczura, W. P., Stiles, M. E. and Vederas, J. C. (1993) *Biochemistry* 32, 310-318.

Sander, C. and Schneider, R. (1991) *Proteins* 9, 56-58.

Shaka, A. J. and Freeman, R. (1983) *J. Magn. Reson.* 51, 161-169.

Storris, R.W., Truckese, D. and Wemmer, D. E. (1992) *Biopolymers* 32, 1695-1702.

Thornton, K. and Gorenstein, D. G. (1994) *Biochemistry* 33, 3532-3539.

Van Belkum, M. J. and Stiles, M. E. (1995) *Appl. Environ. Microb.* 61, 3573-3579.

Vandenbergh, P. A. (1993) *FEMS Microbiol Rev.* 12, 221-238.

Volkman, B. F. and Wemmer, D. E. (1997) *Biopolymers* 41, 451-460.

Vriend, D. (1990) *J. Mol. Graphics* 8, 52-56.

Vuister, G. W. and Bax, A. (1993) *J. Am. Chem. Soc.* 115, 7772-7777.

Wang, J., Hodges, R. S. and Sykes, B. D. (1995) *I*nt. J. Pept. Prot. Res. 45, 471-481.

Wang, Y., Nip, A. M. and Wishart, D. S. (1997) *J. Biomol. NMR* . 10, 373-382.

Waterhous, D. V. and Johnson, Jr. W. C. (1994) *Biochemistry* 33, 2121-2128.

Williamson, M. P., Asakura, T., Nakamura, E. and Demura, M. (1992) *J. Biomol. NMR* 2, 83-96.

Williamson, M. P., Kikuchi, J. and Asakura, T. (1995) *J. Mol. Biol.* 247, 541-546.

Williamson, M.P. (1990) *Biopolymers 29*, 1423-1431.

Wishart, D. S. and Nip, A.M. (1998) *Biochem. Cell Biol.* 76, 153-163.

Wishart, D. S., Sykes, B. D. and Richards, F. M. (1991) *J. Mol. Biol.* 222, 311-322.

Wishart, D. S., Sykes, B. D. and Richards, F. M. (1992) *Biochemistry* 31, 1647-1653

Wishart, D. S., Bigam, D.G., Holm, A., Hodges, R. S. and Sykes, B. D. (1995) *J. Biolmol. NMR.* 5, 67-81.

Wishart, D.S., Willard, L. and Sykes, B.D. (1995) VADAR 1.1 - University of Alberta (ftp://redpoll.pharmacy.ualberta.ca)

Wishart, D. S., Fortin, S., Woloschuk, D. R., Wong, W., Rosborough, T., Van Domselaar, G. V., Schaeffer, J. and Szafron, D. (1997) *Comput. Appl. Biosci.* 13, 561-562.

Wishart, D. S., Kondejewki, L. H., Semchuk, P. D., Kay, C. M., Hodges, R. S. and Sykes, B. D. (1995)*Techniques in Protein Chemistry. VI*, 451-457. AP academic press.

Wuthrich, K. (1986) *NMR of Proteins and Nucleic Acids.* Wiley-Inter-Science, New York.

Yi, G. S., Choi, B. S. and Kim, H. (1994) *Biophysical Journal* 66, 1604-1611.

Young, J. K., Hick, R. P., Wright, G. E. and Morrison, T. G. (1998) *Virology* 243, 21-31.

Zhou, N. E., Kay, C. M., Sykes, B. D. and Hodges, R. S. (1993) *Biochemistry* 32, 6190-6197.

Zhong, L., and Johnson, W. C. J . (1992) *Proc. Natl. Acad. Sci. U.S.A* 89, 4462-4465.

Table 5.1. Redox reaction rates for homologous LeuA and CbnB2 hexapeptides

Hexapeptides	$K_{red}(min^{-1}\mu M^{-1})$ *	$K_{ox}X10^{3}(min^{-1}\mu M^{-1})$ **
#1 AcCSKTKC	0.7	3.6
#2 AcCTKSGC	1.1	1.0
#3 AcCSKTGC	0.4	3.0
#4 AcCSGTKC	0.5	1.3
#5 AcCSGEKC	0.1	1.1
#6 AcCSKEKC	0.3	5.1

*K_{red} - reduction constant with excess (~5x) DTT at pH 5.0-5.5

**K_{ox} - oxidization constant with excess oxygen at pH 8.0-8.5

Table 5.2. ^{1}H NMR chemical shift assignments^a, $^3J_{HNH\alpha}$ coupling constants and HN temperature coefficients for Carnobacteriocin B2 in TFE/H$_2$O, 90%:10% v/v.

Amino Acids	HN	Hα	Hββ	Other H	$^3J_{HNH\alpha}$ (ppm)	HN temp. Coeff. (ppb/°C)
1 Val	--	3.68	2.07	γCH$_2$ 0.89, 0.85		
2 Asn	8.14	5.03	2.89, 2.73	γNH$_2$ 7.17, 6.47	6.7	5.0
3 Tyr	8.25	4.60	3.15, 3.05	2,6 H 7.09, 3,5H 6.77	7.0	10.9
4 Gly	8.21	3.95, 3.86				
5 Asn	7.92	4.71	3.03, 2.82	γNH$_2$ 6.98, 6.14	6.5	2.6
6 Gly	8.16	4.10, 3.79				
7 Val	7.76	4.16	2.21	γCH$_2$ 1.00, 0.98	8.0	1.4
8 Ser	8.44	4.20	4.04, 3.82			
9 Cys	8.18	4.82	2.90, 2.75			2.8
10 Ser	8.19	4.56	4.22	3.98		
11 Lys	8.05	4.23	1.99	γ, δ CH 1.52, 1.74, 1.62, εCH 3.00, εNH 7.43	6.5	7.3
12 Thr	7.84	4.49	4.34	γCH$_3$ 1.23	8.0	1.8
13 Lys	7.91	4.42	1.94	γ, δ CH 1.45, 1.71 εCH 3.00, εNH 7.43	7.0	
14 Cys	7.94	4.80	3.68;3.00			
15 Ser	7.84	4.48	3.92, 3.83			
16 Val	7.49	4.18	1.70	γCH$_2$ 0.84, 0.53	6.3	1.7
17 Asn	8.16	4.75	2.99, 2.74	γNH$_2$ 7.29, 6.37		
18 Trp	7.68	4.44	3.38, 3.25	NH 9.34, 4H 7.52, 7H 7.41, 6H 7.25, 5H 7.16, 2H 7.06	6.5	0.1
19 Gly	8.11	3.95				4.4
20 Gln	7.77	4.16	2.14	γCH$_2$ 2.45, 2.37 δNH$_2$ 6.89, 6.31	7.0	1.0
21 Ala	7.95	4.10	1.51		5.5	4.4
22 Phe	8.39	4.26	3.21	2,6 H 7.15, 3,5H 7.24	6.3	8.2
23 Gln	8.30	4.03	2.35, 2.28	γCH$_2$ 2.45, 2.37 δNH$_2$ 6.89, 6.31	5.5	1.5
24 Glu	8.53	4.03	2.35, 2.28	γCH$_2$ 2.75, 2.52	5.8	5.0

25 Arg	8.02	4.04	1.95	γCH_2 1.80, 1.69 δCH_2 3.21, 3.13 εNH 6.90	5.5	3.4
26 Tyr	8.77	4.21	2.97, 2.83	2,6 H 7.03, 3,5H 6.97	4.9	10.2
27 Thr	8.17	4.21	3.91	1.37		3.7
28 Ala	8.27	4.18	1.56		5.2	4.2
29 Gly	8.19	3.98, 3.87				2.4
30 Ile	8.12	3.95	1.94	γCH_2 1.48, 1.45 δCH_2 0.99, 0.75	5.3	1.1
31 Asn	8.45	4.50	2.91, 2.76	γNH_2 7.21, 6.28	5.3	4.4
32 Ser	8.44	4.19	4.04, 3.82		5.3	3.4
33 Phe	8.11	4.38	3.35, 3.30	2,6 H 7.21, 3,5H 7.27	5.3	2.9
34 Val	8.84	3.54	2.19	γCH_2 1.25, 1.03	5,5	7.4
35 Ser	8.19	4.19	4.05, 3.97			2.9
36 Gly	7.91	3.94, 3.81				1.0
37 Val	7.98	3.66	2.05	γCH_2 0.89, 0.73	5.6	0.4
38 Ala	8.67	4.08	1.48		4.8	6.8
39 Ser	7.95	4.33	4.10, 4.03		5.2	3.6
40 Gly	7.82	4.01				0.1
41 Ala	8.38	4.12	1.46		4.4	5.3
42 Gly	8.20	3.95, 3.85				4.6
43 Ser	7.91	4.41	4.04, 3.92		5.2	2.6
44 Ile	7.63	4.01	1.96	γCH_2 1.64, 1.26 δCH_3 0.98		1.6
45 Gly	8.08	4.00, 3.92				6.3
46 Arg	7.67	4.43	1.94, 1.81	γCH_2 1.67, 1.63 δCH_2 3.20 εNH 6.90	5.7	3.5
47 Arg	7.68	4.47	1.93	γCH_2 1.79, 1.73 δCH_2 3.23 εNH 6.91	5.4	8.5
48 Pro		4.47	2,31, 2.05	γCH_2 2.09 δCH_2 3.78, 3.64		

[a] Chemical shifts are reported in ppm relative to the protonated TFE at 3.88 ppm downfield from DSS.

Table 5.3. Structural and stereochemical statistics for CbnB2 and LeuA[a]

A. Structural statistics		
A1. RMS deviation	CbnB2 (SA)	LeuA (SA)
rmsd from exptl distance restraints (Å)	0.06 ±0.01	0.05±0.02
rmsd from exptl dihedral restraints (deg)	0.21±0.12	0.12±0.09
rmsd from $^3J_{HNN\alpha}$ restraints (Hz)	0.13±0.06	0.21±0.10
rmsd from ^{1}H chem. shift restraints (ppm)	0.23±0.10	0.32±0.12
A2. Deviations from idealized covalent geometry		
bonds (Å)	0.0051±0.0010	0.0043±0.0011
angles (deg)	0.48±0.14	0.81±0.09
impropers (deg)	0.62±0.06	0.76±0.12
overall G-Factor	0.0±0.01	-0.1±0.03
ϕ/ψ in the most favorable region (%)	96.7±1.2	74.7±2.4
ϕ/ψ in additionally allowed region (%)	3.3±1.5	18.2±1.7
ϕ/ψ in generously allowed region (%)	0.0	1.6±2.1
ϕ/ψ in disallowed region (%)	0.0	5.6±1.1
H-bond energy (kcalmol^{-1})	-1.5±1.0	-1.4±1.2
Van der Waals energy (kcalmol^{-1})	-83.5±13.9	-44±15.4
number of bad contacts/100 residues	1.1±0. 8	1.6±1.2

B. Atomic coordinate rms deviation (Å)[c]				
	CbnB2 (residue 19 –39)		LeuA (residue 2-31)	
	backbone	heavy	backbone	heavy
<SA> vs $(SA)_{mean}$	0.48 ± 0.08	0.94 ± 0.11	0.67 ± 0.11	1.04± 0.12

[a] The notation of the structures is as follows: <SA> are the final NMR simulated annealing structures; $(SA)_{mean}$ is the mean structure obtained by averaging the coordinates of the individual SA structures best fit to each other.

[b] Analyzed using PROCHECK-NMR version 3.4.4 (Laskowski et al., 1996).

[c] Analyzed using MolMol version 4.0 (Koradi et al., 1996).

Table 5.4. Average atomic rmsd to their mean coordinates for CbnB2[a]

Residues	Backbone atoms (Å)	Heavy atoms (Å)
19 -39	0.48	0.94
2 - 18	3.51	4.61
9 - 14	1.14	2.18
2 - 7	0.99	1.91

[a] Analyzed using MolMol version 4.0 (Koradi et al., 1996).

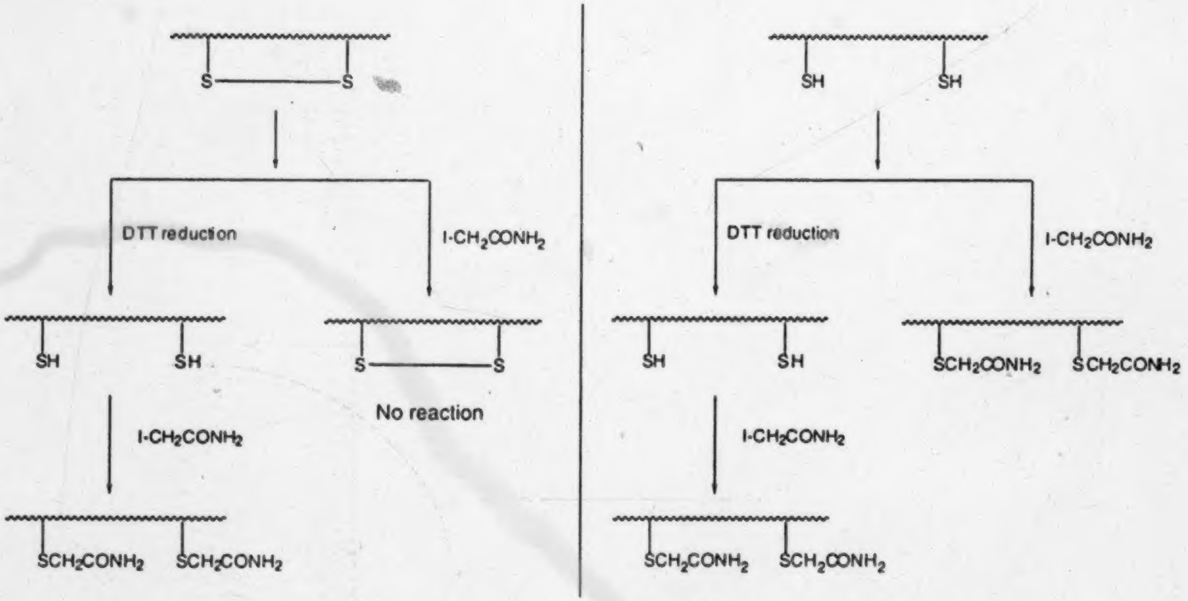

S
S
DTT reduction
I-CH_2CONH_2
SH
SH
I-CH_2CONH_2
SCH_2CONH_2
SCH_2CONH_2
S
S
No reaction
SH
SH
DTT reduction
I-CH_2CONH_2
SH
SH
I-CH_2CONH_2
SCH_2CONH_2
SCH_2CONH_2
SCH_2CONH_2
SCH_2CONH_2

Figure 5.1. Illustration of carboamidomethylation experiments used to investigate the redox status of CbnB2.

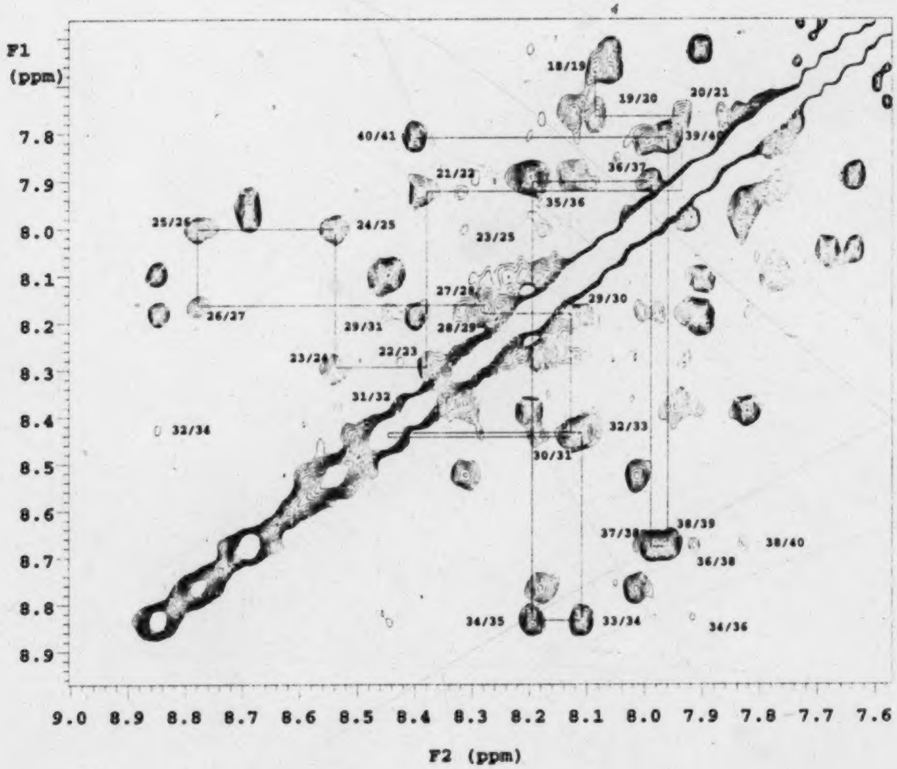


Figure 5.2. The HN-HN region NOESY spectra of CbnB2. Sequential and medium-range NOE connectivities in the helical region (from Trp18 to Ser39) are indicated.

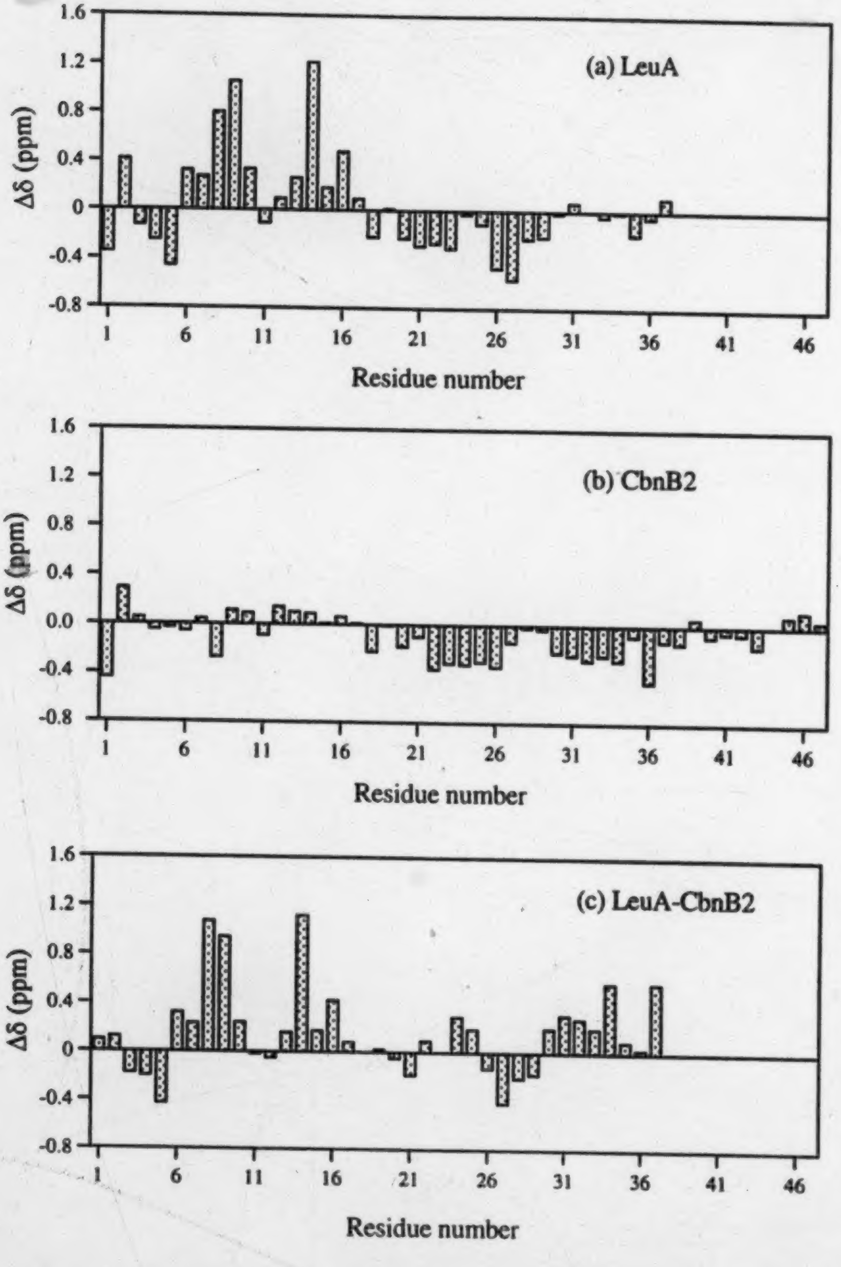
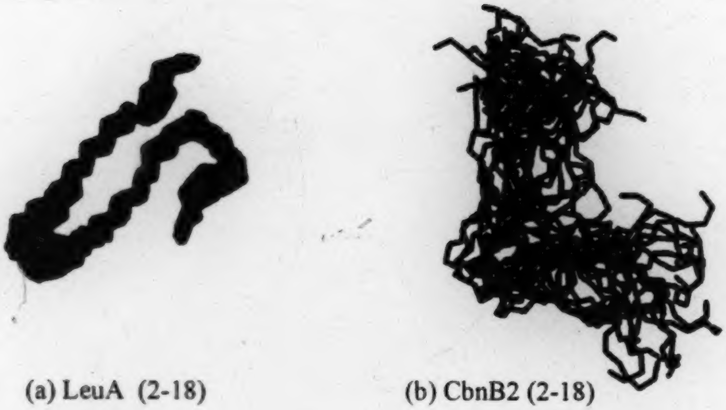
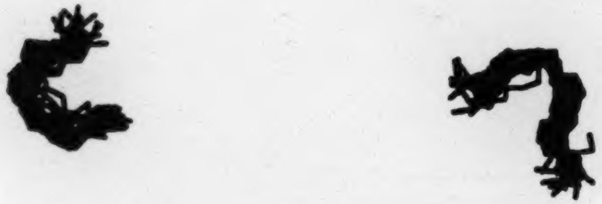


Figure 5.3. Net structural α–proton chemical shifts (observed chemical shift minus the random coil chemical shift (Wishart and Nip, 1998)) for (a) LeuA, (b) CbnB2, and (c) difference between LeuA and CbnB2.



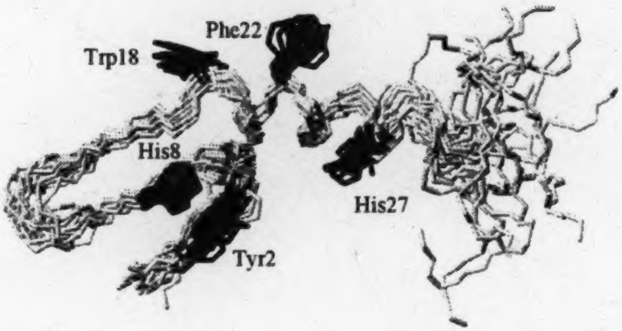
(a) LeuA (2-18) (b) CbnB2 (2-18)



(c) CbnB2 (9-14) (d) CbnB2 (2-7)

Figure 5.4. Backbone superimposition of (a) LeuA (residues 2-18), (b) CbnB2 (residues 2-18), (c) CbnB2 (residues 9-14), and (d) CbnB2 (residues 2-7).

(a) before refinement


Phe22
Trp18
His8
His27
Tyr2

(b) after refinement

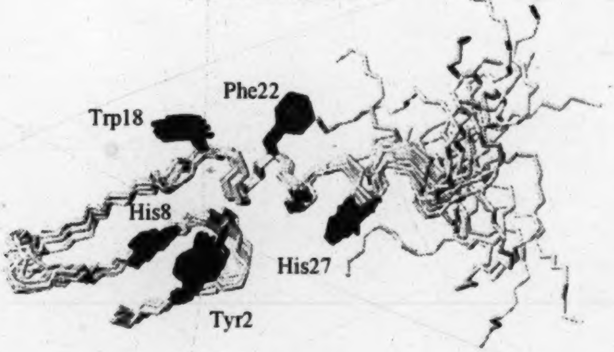

Phe22
Trp18
His8
His27
Tyr2

Figure 5.5. Best-fit backbone atom superimposition of the (a) unrefined and (b) refined LeuA structure.

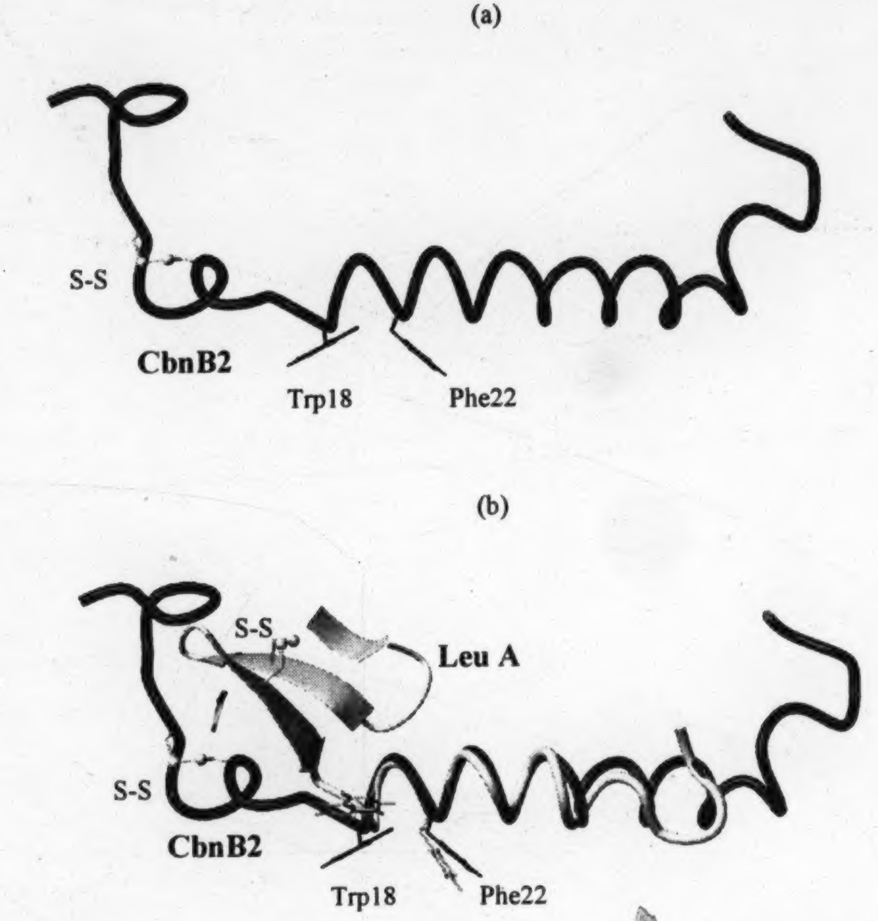

(a)
S-S
CbnB2
Trp18
Phe22
(b)
S-S
Leu A
S-S
CbnB2
Trp18
Phe22

Figure 5.6. Three-dimensional structure of CbnB2 (a) and comparison of CbnB2 and LeuA (b). The superimpostion is based on the alignment of backbone atoms from Trp18 to Phe22 for these two molecules.

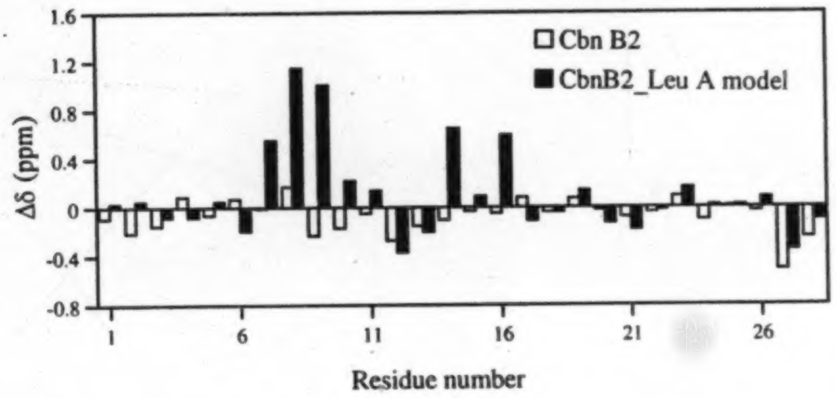


Figure 5.7. The α–proton chemical shift difference between the predicted and observed values for CbnB2.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Leu A	K	Y	Y	G	N	G	V	H	C	T	K	S	G	C	S	V
	✱	✱	|	|	|	|	|	✱	|	•	|	•	✱	|	|	|
Cbn B2	V	N	Y	G	N	G	V	S	C	S	K	T	K	C	S	V

	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
Leu A	N	W	G	E	A	F	S	A	G	V	H	R	L	A	N	G
	|	|	|	•	|	|	✱	✱	✱	✱	✱	✱	✱	•	|	✱
Cbn B2	N	W	G	Q	A	F	Q	E	A	Y	T	A	G	I	N	S

	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48
Leu A	G	N	G	F	W											
	✱	✱	✱	✱	✱											
Cbn B2	F	V	S	G	V	A	S	G	A	G	S	I	G	R	R	P

Figure 5.8. Sequence alignment of CbnB2 and LeuA. The " |" indicates exact residue identity, "." indicates residue similarity, and "★" indicates unlike amino acids.

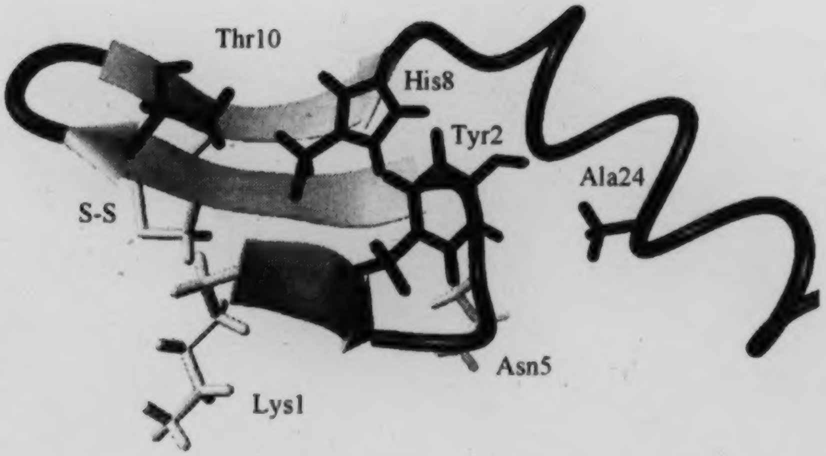

Thr10
His8
Tyr2
Ala24
S-S
Asn5
Lys1

Figure 5.9. Amphipathic β-sheet in the N-terminal of LeuA where Tyr2, His8, Thr10 and Ala24 form a hydrophobic cluster; Lys1, Asn5, Cys9 and Cys14 forms a weak hydrophilic cluster.

Chapter 6

Expression and Characterization of a Single-Chain Anti-PSA Antibody Fragment, scFvB80 Produced in *P. pastoris*

6.1. Introduction

Since the early 1980's it has been known that high serum prostate specific antigen (PSA) levels are well correlated with the presence of prostate cancer (Lilja et al., 1991; Seregni et al., 1996). By using simple immunometric assays it is now possible to quantitate serum PSA levels in a matter of hours (Brawer, 1991; Graves, 1993; Ploch and Brawer, 1994). However, current problems with many commercial tests point to the lack of understanding regarding how PSA interacts with commercially available antibodies (Wu, 1994; Luderer et al., 1995; Irani et al., 1996). In an effort to better understand the interaction between PSA and diagnostic antibodies that are directed against it, we have designed and cloned a single chain fragment (scFvB80) of an anti-PSA antibody (B80.3).

B80.3 is a murine monoclonal IgG first developed by Biomira Inc (Edmonton) in 1992. Originally developed as a potential metastatic bone cancer imaging reagent (Jette et al., 1996), B80.3 has very high affinity ($K_d=1x10^{-9}$ M) for both free and α-antichmostrypsin complexed PSA. More importantly, its gene sequence is known--making it one of only two anti-PSA antibodies that has been fully cloned and sequenced. Based on earlier epitope scanning experiments (Jette et al., 1996; Jette, 1997) along with homology modeling and detailed sequence comparisons (Jette et al., 1996; Bridon and Dowell, 1995; Villoutreix et al., 1996) it appears that B80.3 binds to a surface-exposed loop in PSA between residues 50-59. Interestingly this loop is not masked by α-antichmostrypsin - a protein commonly complexed with PSA. Furthermore, the 50-59 loop is not conserved between human glandular kallikrein (a close homologue of PSA) and PSA, itself. This makes B80.3 one of only handful antibodies that are truly specific to PSA. Recent epitope scanning experiments conducted on a panel of 25 Mabs directed against PSA and hK2 support these conclusions (Piironen et al., 1998).

We have previously reported on the design and construction of a single chain variant of B80, consisting of the two B80.3 variable domains (V_L and V_H) joined together by an artificial peptide linker. This construct was cloned into specially designed bacterial expression system (Zhang et al., 1997; Zhang et al., submitted). The small size of such an antibody fragment, combined with its amenability to genetic manipulation suggested that high-level production of scFvB80 would permit detailed X-ray crystallographic and NMR studies of both the antibody fragment and its protein/peptide substrates (Freund et al., 1993 and 1994; Zdanov et al., 1994; Raag and Whitlow, 1995; Malby et al., 1998). However, despite our best efforts at optimizing scFvB80 production in *E. coli*, we were unable to improve the yields to levels sufficient for detailed structural characterization. Indeed, this negative result is quite typical of attempts to produce single chain antibodies using prokaryotic expression systems (Stader and Silhavy, 1990; Kipriyanov et al., 1997).

Recently, it has been reported that scFv production can be improved by expressing these slow-folding eukaryotic proteins in the methylotrophic yeast -- *Pichia pastoris* (Luo et al., 1995 and 1996; Ridder et al., 1995; Eldin et al., 1997). Based on these studies we moved our scFvB80 gene into *P. pastoris* and assessed its potential as an expression vehicle. In this report we wish to describe the successful cloning, high level production (>15 mg/L) and rapid purification of this single chain antibody from *P. pastoris*. Competitive ELISA's conducted on scFvB80 indicate this antibody fragment exhibits ~20% of the activity of the full-length B80.3 antibody. We also show that scFvB80 can be labeled with ^{15}N using a novel, inexpensive, high yielding method designed specifically for *P. pastoris* expression. Additional NMR studies conducted on this scFvB80 fragment and its peptide mimetopes (mimics of the PSA epitope) along with computer-aided ligand docking simulations have allowed us to create a working model of the scFvB80/PSA complex. This model, along with the high-level expression system reported here, should lay the groundwork to perform more detailed structural studies of this clinically important antibody.

6.2. Materials and Methods

Cells, Antibodies and Growth Media. The *Pichia pastoris* strain G115(his4) was purchased from Invitrogen (San Diego, CA). Antibodies used in ELISA and

Western Blot experiments were obtained from three sources. The full-length Mab B80.3 was a gift from Biomira Inc. (Edmonton, Canada). The anti-c-myc antibody 9E10 (Evan et al., 1985) was produced in-house from a clone obtained from the American Type Culture Collection (ATCC, New Haven, CT USA). The bi-specific B80.3/HRPO Mab and purified prostate-specific antigen were gifts from the laboratory of Dr. M.R. Suresh (University of Alberta). *P. pastoris* cells were cultured at 30 °C on buffered minimal glycerol medium (BMGY), containing (per litre) 10 g of yeast extract, 20 g peptone, 100 mL of 1M potassium phosphate buffer (pH 6.0), 100 mL of 13% (w/v) yeast nitrogen base, 2 mL of 0.02% biotin (w/v) and 100 mL of 10% (v/v) glycerol. Protein expression was induced in the recombinant *P. pastoris* cells cultured with buffered minimal methanol medium (BMMY), which is identical to BMGY except that glycerol is replaced by 5% methanol.

Construction of the scFvB80 Expression Plasmid. As described previously (Zhang et al., 1997, Zhang et al., submitted), the gene encoding the scFvB80 contains the B80.3 V_H domain tandemly linked by a synthetic linker peptide $(GGSGS)_3$ to the B80.3 V_L domain. At the C-terminal end of this construct is an 11-residue c-myc peptide epitope (Munro & Pelham, 1986) to facilitate immunodetection (Figure 6.1). To prepare the *Pichia* expression plasmid, the DNA fragment containing both the scFvB80 gene and the c-myc tag was cut from the original bacterial construct using PvuII and HindIII. The fragment was then gel purified, end-filled with Klenow fragment and dephosphorylated. This blunt-ended fragment was ligated into a pPIC9 expression vector, which had been previously linearized with SnaB1/EcoRI and end-filled. The pPIC9 vector contains the α-mating factor signal sequence to facilitate secretion and the HIS4 gene for selection of recombinant yeast clones.

Transformation and Screening of *P. pastoris* cells. *P. pastoris* cells were transformed with linearized DNA using electroporation. The *Pichia* expression plasmid described above was linearized with BglII. Competent cells were prepared as follows: 500 mL of GS115 cells were grown at 30 °C overnight to an A_{600} (UV absorbance at 600 nm) of 1.3 to 1.5. The cells were then centrifuged at 1500 x g for 15 minutes and the pellet resuspended in 500 mL of ice cold sterile water. The process was repeated, first with 250 mL of water, then with 20 mL of ice cold sorbitol (1.0 M) and finally with 1.0 mL of ice cold sorbitol (1.0 M) to give a final volume of 1.5 mL of competent cells. A small aliquot (40 μL) of competent cells was then mixed with 1.0 μg of linearized DNA and transferred to an ice cold 2.0 mm electroporation cuvette. The cell-DNA mixture was incubated on ice for five minutes and then pulsed at 1800 volts for 5.0 ms using a model EC100 electroporator. The electroporated cells were transferred into a culture tube with the addition of 1.0 mL of ice cold YPD (1% yeast extract / 2% peptone / 2% dextrose) media and incubated for one hour at 30 °C with gentle shaking. Cells were then plated onto MD (Minimal Dextrose containing per liter 1.34 g yeast nitrogen base, 2 ml 0.2% biotin, 20 g dextrose) agar plates and incubated at 30 °C for two days. Colonies which grew on the selective MD agar plates were screened for AOX-1 disrupted transformants by selecting those cells which grew on MD but not on MM (Minimal Methanol containing per liter 1.34 g yeast nitrogen base, 2 ml 0.2% biotin, ml methanol) plates. Each of the positive clones (10 in total) was inoculated into 10 mL of growth media (BMGY) and incubated for two days at 30 °C with vigorous shaking. The cells were harvested by centrifugation (1500 x g for 10 minutes), resuspended in BMMY (to induce scFvB80 expression) and re-incubated for 24 hours at 30 °C with vigorous shaking. Secretion of scFvB80 was examined by SDS-PAGE of the culture supernatants and confirmed by Western Blotting using the 9E10 anti-c-myc antibody. The highest expressing clone was selected for further scale-up and production.

Expression and Purification of scFvB80. A 10 mL overnight preculture of the best expressing *P. pasotris* clone was used to inoculate one liter of growth media (BMGY). The culture was incubated at 30 °C for around two days with vigorous shaking until the A_{600} was between 40 and 50. The cells were then harvested by centrifugation (1500 x g for 10 minutes at 4 °C), resuspended in 200 mL of expression media (BMMY) and incubated at 30 °C with vigorous shaking for 18 to 24 hours. Upon completion, cells were removed from the media by centrifugation (1500 x g) for 10 minutes at 4 °C. A three-step purification protocol was used to isolate secreted scFvB80 from the cell supernatant. Solid ammonium sulfate was added, with stirring, to the supernatant to bring the salt concentration to 60% of saturation at 0 °C. The pH was adjusted to 7.5 using 10 N NaOH and the solution left on ice for 15 minutes. The resulting protein precipitate was separated from the medium by centrifugation (10,000 rpm (Sorvall Instruments, GS-3 rotor) for 10 minutes at 4 °C). Further purification and concentration of the scFvB80 was achieved using T-Gel affinity chromatography (Belew et al., 1987). 10 mL of T-gel (Pierce, Rockford, IL, USA) was placed in a 22 x 40 mm column previously equilibrated with four column volumes of 1.0 M $(NH_4)_2SO_4$, 50 mM $NaHPO_4$, pH 7.5 loading buffer. The protein precipitate was dissolved with 10 mL above loading buffer and then slowly loaded onto the column. After washing the column with 40 mL of loading buffer, the bound scFvB80 was eluted with 50 mM $NaHPO_4$ pH 7.5 buffer. The T-gel column can be reused a number of times before requiring regeneration with 8.0 M guanidinium chloride. To remove the remaining colored contaminants, those fractions containing scFvB80 were slurred with 10 mL of DE52 anion exchange resin and gently filtered. The filtrate was collected and stored in a freezer until further use. Throughout the purification and storage of scFvB80, 1.0 mM EDTA was added to all solutions to prevent proteolytic degradation of the antibody fragment.

Isotopic Labeling of scFvB80. Repeated attempts at labeling scFvB80 with ^{15}N enriched minimal media using methods previously described (Laroche et al., 1994; Loewen et al., 1997) were largely unsuccessful. This led us to develop a novel, recyclable labeling approach that uses previously harvested *P. pastoris* cells as feedstock for improved *P. pastoris* growth and concomitant protein expression. To prepare the ^{15}N feedstock (yeast extract) *P. pastoris* cells were grown (for 48 hours) in ^{15}N enriched minimal media, containing (per liter) 5g $(^{15}NH_4)_2SO_4$, 12g KH_2PO_4, 4.7g $MgSO_4.7H_2O$, 0.36g $CaCl_2.2H_2O$, plus the following trace elements: 0.2µM $CaSO_4.5H_2O$, 1.25µM KI, 4.5µM $MnSO_4.4H_2O$, 2.0µM $Na_2MoO_4.2H_2O$, 0.75µM H_3BO_3, 17.5µM $ZnSO_4.7H_2O$, and 44.5µM $FeCl_3.6H_2O$. The cells were then spun down (1500 x g for 10 minutes), resuspended 500 ml H_2O and boiled for 20 minutes to release the cell contents. After allowing the solution to cool to 25 °C, the pH was adjusted to 7.0 and 0.1 mg/mL of pronase (Boehringer, Basel, Switzerland) was added. The suspension was then incubated at 37 °C for 3 hours to permit complete proteolytic digestion. After centrifuging (1500 x g for 10 minutes) to remove cellular debris, the supernatant ("homemade yeast extract") was autoclaved and stored. To produce ^{15}N labeled scFvB80, *P. pastoris* cells containing the scFvB80 gene were grown in 1.0 L of "rich" growth media, which contains (per liter) 500 mL of the ^{15}N enriched "homemade yeast extract", 100 mL glycerol (10%

v/v), 100 mL 1M pH 6.0 KPO_4 buffer, 3 g yeast nitrogen base (without amino acids or ammonium sulfate), 2 mL of 0.02% biotin (w/v), and 2.0g $(^{15}NH_4)_2SO_4$. Cells were induced in 200 mL of "rich" induction media, which is essentially identical to the above growth media except that glycerol was replaced by methanol (5% v/v). After induction and harvesting, the cells would be recycled and the entire process repeated as before. ^{15}N labeled protein was isolated, purified and concentrated using the same protocols described earlier.

Biochemical Characterization of scFvB80. SDS-PAGE, Western Blotting and Competitive ELISAs were used to characterize purified scFvB80. Samples were subjected to electrophoresis on 15% (w/v) denaturing polyacrylamide gels and stained with Coomassie Blue G200. Molecular weight markers were used to confirm the size (28 kDa) of the B80 antibody fragment. As noted earlier, the scFvB80 protein expressed in this system was tagged with a c-myc peptide epitope. This allowed Western blotting to be used to confirm the correct translation (and identity) of the scFvB80 gene product. Western blotting was performed as follows: after SDS-PAGE electrophoresis, the protein was transferred electrophoretically onto a NitroPure nitrocellulose transfer membrane (Amersham, NJ, USA). The membrane was blocked with 5% skim milk in PBS (phosphate-buffered saline) at room temperature for two hours. Anti-c-myc IgG antibody (2.0 μg/mL final concentration) was incubated with the membrane at 4 °C for 16 hours. After washing three times with PBS/0.02% Tween-20, bound antibody was detected using horseradish peroxidase conjugated anti-mouse IgG and the Enhanced Chemiluminescence Detection kit (Amersham, NJ, USA).

A competitive ELISA was used to determine the relative affinity of scFvB80 for PSA. Anti-PSA antibody, B87.3, was coated on a Maxisorb microtitre plate (Nunc) at a concentration of 1.0 μg/well. The plate was blocked with 1% BSA (bovine serum albumin) in PBS for one hour at room temperature. 50μL of scFvB80 (5 μg/ml) or standard, 30μL of a 1/1000 dilution of mouse ascities containing anti-PSA bispecific B80.3/HRPO antibody with 5.0 μg/mL of HRPO, and 30μL of 50 ng/mL PSA was added to each well. The plate was incubated a room temperature with shaking for 15 minutes. After washing three times with a PBS/0.02% Tween 20 solution, immobilized antibody was detected using tetramethylbenzidine (TMB) as a two-part substrate (BioRad). After 15 minutes the color reaction was stopped with 1.0 M phosphoric acid. The absorbance of each well was measured at 450 nm using a Vmax microplate reader (Molecular Dynamics).

Peptide Synthesis. Two peptides, TLGRPSGLVG (a PSA epitope analog) and WGFDFDFGS (a PSA mimetope analog) were prepared to investigate scFvB80-ligand interactions (Jette, 1997). A third peptide (SRLKNFVRE), having no homology to these two, was also synthesized to serve as a non-interacting control. All three peptides were assembled on a Biosearch 9500 automated peptide synthesizer, modified to accommodate Fmoc/HBTU chemistry. Each peptide was assembled on Rink-MBHA resin (ca. 260mg; 0.77 meq/g; 0.2 mmol) using automated cycles of deprotection (DMF/20% piperdine), activation (HBTU/DMF/NMM) and coupling (in DMF with N_2 bubbling). The peptides were cleaved with 10 mL of a cleavage cocktail: TFA: H_2O: DTT (95:2.5:2.5), precipitated with cold ether (30 mL), washed (cold ether) and lyophilized. Crude peptides were purified by HPLC (Waters - System 501) on a reversed-phase C-8 21x25 mm Zorbax 300 s.b. column using a binary gradient of aqueous 0.1% TFA (solvent A) and acetonitrile containing 0.1% TFA (solvent B) at a flow rate of 8.0 mL/min. The purity of each peptide was confirmed by analytical RP-HPLC (Beckman System Gold), electrospray mass spectrometry (Fisons VG Trio 2000 ES-MS) and NMR spectroscopy.

NMR of scFvB80. For NMR studies, scFvB80 samples were concentrated to ~0.5 mM using Centraprep-10 centrifugal concentration units (Amicon) in a buffer (pH 6.0) containing 50 mM KH_2PO_4, 150 mM NaCl and 1 mM EDTA. D_2O (10% v/v) was added to act as a lock signal. All ^1H NMR experiments were performed on Varian Unity 500 or 600 MHz NMR spectrometers equipped with 5-mm triple resonance probes at a temperature of 25 ± 0.1 °C. 1-D NMR experiments were conducted using a sweep width of 7000 Hz and a relaxation delay of 2.5 seconds. Water suppression was achieved using solvent presaturation during the relaxation delay period. ^1H-^{15}N HSQC (Kay et al., 1992) experiments were typically performed with an ^{15}N sweep width of 2000 Hz and an ^1H sweep width of 6000 Hz. A total of 2048 complex points were collected along the t_2 domain (^1H) and 128 increments along the t_1 domain (^{15}N). The relaxation delay was 2.5 s and the refocusing delay 4.9 ms. Data in both dimensions were zero-filled to create a 4K x

4K data set which was further processed using a shifted sine-bell square weighting function. All NMR spectra were referenced (directly or indirectly) to internal DSS (Wishart et al., 1995).

Peptide Interactions with scFvB80. All scFvB80 samples were prepared as described previously. The pH for each sample was adjusted to 5.80 ± 0.02 by adding small aliquots of 0.1 M KOH/HCl. A ^{1}H spectrum of the protein (alone) was first collected using a sufficient number of scans (1024) to provide a very high signal-to-noise ratio. After the "protein only" spectrum was collected, 50 μL of peptide stock solution (1.0 mg peptide in 200 μL 80%/20% H_2O/D_2O, 50 mM phosphate, pH 5.80 ± 0.02) was added and the pH checked and readjusted to 5.80 ± 0.02 as necessary. The molar ratio between scFvB80 and the added peptide was approximately 1:1. A second ^{1}H NMR spectrum (peptide + protein) was then collected using the same spectral collection conditions as before. A difference spectrum was obtained by subtracting the "protein only" spectra from the "protein + peptide" spectra. Finally a third ^{1}H NMR spectrum (peptide only) was collected, again using identical conditions as above (concentration, buffer and pH). The peptide interactions with scFvB80 were evaluated by comparing the "peptide only" spectra with the difference spectra. All data processing and spectral subtraction were performed using Varian VNMR (version 5.0) software.

To facilitate the peptide interaction studies, all three peptides (without scFvB80) were assigned using conventional 2-D NMR techniques (Wuthrich, 1986). TOCSY experiments were carried out using standard pulse sequences (Bax and Davis, 1985a) with spin lock (MLEV-17) mixing times ranging from 50 to 100 ms. NOESY (Jeener et al., 1979; Kumar et al., 1980) and ROESY experiments (Bax and Davis, 1985b) were collected with acquisition times set to 0.2 seconds, relaxation delays of 2.0 seconds and mixing times ranging from 150 to 300 ms. Prior to Fourier transformation, the data matrix was zero-filled to 4K x 4K complex points and multiplied by an approximate 90^o-shifted sine-bell squared weighting function in both dimensions. All two-dimensional spectra were collected using 256 t_1 increments and spectral widths of 6000 Hz in both dimensions.

Models of scFvB80 and its Target Interactions. Homologous protein with known structure for scFvB80 were obtained using the NCBI BLAST on-line server (Madden et al., 1996). Three antibody structures with high (>60%) sequence identity were found: 1) the scFv fragment of C219 (PDB#: 1AP2); 2) the V_L domain of neutralizing antibody 59.1 (PDB#: 1ACY); and 3) the V_H domain of IgG2a Kappa (PDB#: 1PLG) were selected to serve as structural templates for scFvB80 and its V_L/V_H domains, respectively. After alignment, the sequences of both target and template proteins together with the coordinates of the template proteins were submitted to the WHATIF on-line server (Vriend, 1990; http://swift.embl-heidelberg.de) which created initial structures of scFvB80 and its V_L and V_H domains. The resulting structures were further refined with 500 steps of energy minimization using X-PLOR (Brunger, 1992). During minimization, the target function contained only quadratic harmonic potential terms for covalent geometry (bonds, angles, planes, and chirality) and a quadratic Van der Waals repulsion term for the nonbonded contacts. All three model structures were evaluated using PROCHECK v3.4.4 (Laskowski et al., 1996).

Automated docking simulations were conducted using DockVision 1.0 (Hart and Read, 1992; Hart et al., 1997). The energy minimized structure of scFvB80 was used as the target molecule while a model structure of PSA (PDB code: 1PFA; Villoutreix et al., 1994) served as the rigid ligand. Atomic interaction grids with 0.5 Å step sizes in a 50-Å^3 box centered on the complementary-determining regions (CDRs) (Chothia and Lesk, 1987; Chothia et al., 1989) of scFvB80 were used to define the search region for the target molecule. A hybrid Monte Carlo/simulated annealing algorithm was used to generate energetically favorable positions for the PSA ligand in the specified search region (Hart et al., 1992). To sample as much of the conformational space as possible, 1000 independent Monte Carlo searches were performed with 100 constant temperature simulated annealing cycles. The force field used in the simulation is a Lennard-Jones 6-12 function together with a standard electrostatic function. After the sampling run was completed, the top 10 energetically favorable binding modes were selected for further analysis. Intermolecular hydrogen bond energies, free energies of binding (Eisenberg and McLachlan, 1986) and electrostatic potentials were used to evaluate the stability of the high-scoring binding modes. During the evaluation of free energy of binding, the accessible surface area were calculated using VADAR (Wishart et al., 1995).

6.3. Results and Discussion

Cloning and Expression of scFvB80 in *P. pastoris*. As described previously our linearized scFvB80-pPIC9 construct was introduced into competent *P. pastoris* cells by electroporation, after which the transformed cells were incubated on the appropriate selection media. Approximately 15% of His+ clones exhibited reduced growth, indicating integration of the scFvB80 gene into the yeast AOX1 site. A total of 10 transformant colonies were selected and cultured for further evaluation. The quantity of secreted scFvB80 in each of the cultured supernatants was monitored by SDS-PAGE. All 10 transformants produced significant amounts of protein (data not shown) with the expected molecular weight of ~28 kDa as indicated by SDS PAGE. The identity of the protein band seen on the gels was confirmed by Western blotting using an anti c-myc antibody which selectively bound the scFvB80 c-myc tag (Figure 6.2). A series of test growths monitored by SDS-PAGE indicated that the optimal induction time for scFvB80 expression in *P. pastoris* is ~24 hours at 30 °C. This differs somewhat from the 48 hours (30 °C) suggested by the manufacturer. We found that the levels of scFvB80 significantly decreased after 24 hours, possibly due to proteolytic degradation. By harvesting at 24 hours we consistently recovered greater than 15 mg/L of purified scFvB80. As shown in Figure 6.2, scFvB80 is the only protein that can be detected by SDS-PAGE.

Purification of scFvB80. To purify scFvB80 we devised an efficient three-step protocol that consistently yields >99% pure protein that is essentially free of small molecule contaminants. The protocol begins with an ammonium sulfate precipitation step followed by a thiophilic affinity column and concludes with a DE52 ion exchange "filtration" step. Thiophilic or thio-affinity matrices (T-gel) were first used to purify single chain antibody fragments by Schulze et al. (1994). We found the T-gel to be particularly useful at removing most of the co-precipitated contaminants generated from the earlier ammonium sulfate precipitation step. An additional advantage of using this inexpensive method is that the protein can be concentrated from large volumes of supernatant. This feature is especially useful for large-scale purification or purification from large solution volumes. As with most T-gel purification protocols, scFvB80 was selectively bound to the thio-affinity matrix using high salt concentrations (1.0 M $(NH_4)_2SO_4$/50 mM $NaHPO_4$, pH 7.5) and eluted from the column at low salt concentrations (50 mM $NaHPO_4$, pH 7.5). Most of the colored contaminants did not bind to the column but eluted in the flow-through fractions. NMR spectroscopy indicated that even after this T-gel purification step, a detectable number of small molecule contaminants still existed in the scFvB80 fractions. Removal of these compounds was achieved by passing the protein solution through a DE52-cellulose column which effectively "filtered" out the last of the small molecule contaminants. The high purity of the final product was demonstrated by SEC-HPLC (data not shown) and by NMR spectroscopy. A one-dimensional ^1H NMR spectrum of purified scFvB80 is shown in Figure 6.3. The downfield α-proton peaks (5-6 ppm) and the upfield side chain proton peaks (0 to -2 ppm) indicate that the protein exhibits a good deal of tertiary structure (i.e. a native fold).

Initially we found that the purified (or partially purified) scFvB80 was highly susceptible to proteolytic degradation. In particular, after an overnight dialysis at 4 °C, most of the protein was typically cleaved by an unknown endogenous protease. In an attempt to stabilize the protein against proteolysis a number of protease inhibitors were tested, including 1 mM PMSF (phenylmethylsulfonyl fluoride), 1 mM EDTA (ethylenediaminetetraacetic acid), 1 μg/mL pepstatin A, 1 μg/mL leupeptin, and 5 μg/mL aprotinin. In the end, 1 mM EDTA was found to be the most effective in preventing scFvB80 degradation. This suggests that a small amount of an unknown metal-protease (which is undetectable by SDS-PAGE or NMR spectroscopy) is being co-purified with our scFvB80. Attempts at altering the purification protocol to eliminate this protease have so far been unsuccessful. However, the addition of EDTA to all solutions used in the subsequent purification and concentration steps has allowed samples of scFvB80 to be kept for several months with no evidence of degradation.

Biochemical Characterization of Purified scFvB80. The binding affinity of purified scFvB80 to PSA was compared to that of the full-length B80.3 Mab using a competitive ELISA (Figure 6.6). The bi-specific antibody used in this assay has one B80.3 binding site that recognizes an epitope on PSA while the other binding site recognizes an epitope on HRPO (Kreutz and Suresh, 1995). The use of this bispecific antibody eliminates the need to either conjugate HRPO to B80.3 or to use a secondary anti-mouse IgG/HRPO conjugate. The relative affinity of Mab B80.3 to PSA, using this method, was determined to be 15 nM. This compares well to the published value for B80.3 of 12 nM (Kreutz and Suresh, 1995; Jette, 1997). Using protein samples prepared as described above, the relative affinity of scFvB80 was found to be 85 nM in this assay (Figure 6.4). This indicates that scFvB80 has approximately 20% of the affinity of the full-length B80.3 Mab. This loss of binding activity was not entirely unexpected as many other scFv's generally exhibit a 5 - 10X reduction in substrate affinity (Milenic et al., 1991; Devlin et al., 1995; Bruyns et al., 1996; Benedict et al., 1997; Nicholson et al., 1997; Luo et al., 1997a and 1997b). It is generally thought that the reduction in substrate affinity may be due to a loss in conformational stability brought on by replacing a substantial part of the full-length antibody with a single peptide linker (Kortt et al., 1994; Alfthan et al., 1995). It has also been suggested that the introduction of a peptide linker can lead to a slightly altered conformation around the binding site and therefore a reduced or modified substrate interaction (Reiter et al., 1994a and 1994b; Brinkmann et al., 1997).

Isotopic Labeling scFvB80 in *P. pastoris*. To date there have been three published protocols describing the isotopic labeling of proteins in *P. pastoris* (Laroche et al., 1994; Loewen et al., 1997; Wood and Komives, 1999). All three methods make use of essentially the same defined, minimal media containing appropriate salts, vitamins and cofactors along with a labeled carbon and/or nitrogen source. The earliest effort by Laroche et al. involved using standard shaker flasks in a simple incubator for cell growth. They found that while the cells grew vigorously, there was a 10x reduction in overall protein yields. Despite the reduction in yield, this was still sufficient for their purposes as the protein expressed at 1.7 g/L in rich BMGY media. Subsequent attempts have shown that yields can be increased if the cells are grown in a carefully optimized, large volume fermentor system (Loewen et al., 1997; Wood and Komives, 1999). Initially we chose to use a minimal media/shaker system (similar to that described by Laroche et al.) to prepare ^{15}N labeled scFvB80. Repeated attempts in our laboratory indicated that while the cells grew vigorously, the protein was not expressed at detectable levels. Evidently the modest expression level (15 mg/L) of scFvB80 was insufficient to overcome the expected 10-20 fold drop in yield. Because of the expense of turning to a large-scale fermentor system and the limited local availability of such a facility, we decided to investigate the possibility of creating a "rich" isotopic medium out of the discarded (but ^{15}N labeled) *P. pastoris* cell paste. We hypothesized that the creation of a rich medium, similar to the BMGY used in standard non-labeled expression, would allow not only more vigorous cell growth but also more vigorous protein production. The simple protocol we devised involves taking *P. pastoris* cells grown in standard shaker flasks to maximal cell density on isotope enriched minimal media (as per Laroche et al., 1994) and then treating them to produce a proteolyzed yeast extract. The treatment, which only requires a few hours, initially involves boiling the cells to release their contents, removing insoluble debris, treating the material with pronase (to digest the now-denatured proteins) and finally autoclaving the resulting product. This "home-made" isotopically labeled yeast extract (supplemented with additional salts and cofactors) essentially serves as a rich-medium substitute for minimal or defined media. As seen in Figure 6.5 cells grown in this "homemade" medium exhibit essentially the same protein yields as those grown in BMGY medium. Tests using two other proteins cloned into *P. pastoris* (an unrelated scFv fragment and a type II antifreeze protein) indicate that the protein yields with this easily-prepared media are comparable to those seen in BMGY and substantially better than those seen in minimal media (data not shown). Using this "homemade" yeast extract and the protocols described previously, we have successfully prepared more than 30 mg of ^{15}N labeled scFvB80. A selected region of ^{15}N HSQC NMR spectrum of the ^{15}N labeled material is shown in Figure 6.6.

We believe this particular protocol for isotopic labeling proteins in *P. pastoris* has two important advantages. First, it is amenable to small-scale, inexpensive shaker flask growths and second, it allows one to continuously recycle expensive isotopic materials. Both of these features can lead to substantial savings in isotope and equipment costs. The fact that one can get high levels of isotopically labeled protein without having to resort to using large, unfamiliar or costly fermentor systems is perhaps the main selling point of this particular protocol. However, because the method makes use of cells and cell products that one normally discards, it also offers the opportunity to recycle expensive isotopes and to minimize the need for additional ^{15}N or ^{13}C supplements. Based on our results to date, we estimate that this home-made media can be recycled up to three times before it is necessary to prepare a fresh batch. Amino acid analysis, elemental analysis and NMR studies of the home-made media indicate that it shares many of the same properties (peptide size, amino acid content, carbon/nitrogen/oxygen/sulfur content) as commercial yeast extract. While the results to date are promising, it must be said that given the subtleties of protein expression it is not likely that this method will work for all proteins under all growth conditions. Modifications to the media composition or preparation may be necessary to adapt this home-made approach to certain hard-to-express proteins. Of course, nothing prevents one from using this "rich" medium in a fermentor system to help boost yields or overcome unexpected expression difficulties.

Peptide Interaction with scFvB80 by NMR. Recently Jette (Jette et al., 1996; Jette, 1997) used epitope scanning and phage panning to identify hexa- and decapeptides, respectively, that bind to MAb B80.3. From Jette's epitope scanning work, one of the best binding hexapeptides was found to have the amino acid sequence: RHSLFH, while one of the more reactive set of decapeptides identified from phage panning had the consensus sequence: TLGRRS[GP]LV. Comparisons to the PSA sequence indicated that the stretch of amino acids: LLGRHSLFHP (residues 50-59) exhibited good similarity to both the hexapeptide and decapeptide. In addition to this work, Jette (1997) also discovered a completely unrelated peptide sequence -- called a mimetope -- that exhibited equally strong affinity to the B80.3 molecule. Its sequence was determined to be WGFDF. In an effort to independently confirm these earlier findings of Jette et al. while at the same time gaining a more detailed molecular understanding of the molecular interactions involved in B80.3-PSA interactions, we decided to investigate the interactions of scFvB80 with these two peptides (TLGRPSGLVG and WGFDFDFGS) along with a control peptide (SRLKNFVRE) using NMR spectroscopy:

It is well known that NMR can offer an exquisitely detailed molecular picture of protein-ligand interactions through such techniques as transferred NOE's, isotope edited NOE's, chemical shift perturbations and difference spectroscopy (Tsang et al., 1992; Campbell et al., 1997; Hubbard et al., 1997; Huang et al., 1998). Because of the limitations of our system (large size of the target, low affinity, lack of labeled peptide or protein) we chose to use simple difference spectroscopy to investigate the peptide/scFvB80 interactions. In difference spectroscopy NMR, the spectra of scFvB80 and the peptides, both free and in complex with scFvB80 are collected separately. After collection and careful scaling one then subtracts the scFvB80 spectrum from the each of the peptide-scFvB80 complexed spectra and looks for perturbations to the difference spectrum relative to the spectrum collected for the corresponding free peptide(s). As seen in Figure 6.7 a striking change in the intensity of the subtracted spectra of both the PSA epitope peptide (TLGRPSGLVG) and the PSA mimetope peptide (WGFDFDFGS) but not for the control peptide (SRLKNFVRE - which has no sequence homology to either the epitope or mimetope peptides) . In particular the intensity of the peptide peaks was found to decrease by nearly 60% for the epitope peptide and more than 80% for the mimetope peptide. This loss in resonance intensity is likely due to the tight binding of ~60-80% of the availalbe peptide molecules to the scFvB80 antibody and the concommitant broadening of the peptide resonances due to their greatly shortened T2 relaxation time. Under the conditions used for this experiment, the complex formed between the scFvB80 antibody and its peptide epitopes was likely close 60 kD, which would lead to peptide linewidths that would be not much above the baseline. The fact that no loss in intensity was seen for the control peptide after spectral subtraction indicates that no peptide complex was formed. This result illustrates that the peptide binding is specific and not due to the formation of non-specific complexes. In addition to theses obvious intensity changes for the epitope and mimetope peptides, some chemical shift changes were evident as well. These chemical shift perturbations appear to arise from a smaller population (20-30%) of the peptides that are in rapid exchange with the antibody (Figures 8 and 9). In particular, chemical shift perturbations can be seen for the Tyr1 ring proton of the mimetope peptide (Figure 6.9), indicating that this side-chain may be strongly interacting with scFvB80. Tests conducted with the control peptide indicate little or no significant chemical shift perturbations or line broadening in the difference spectrum (Figure 6.10). It is important to note, however, that the resonances observed in these difference spectra can arise not only from the peptides but also from the resonances belonging to scFvB80 itself. In general chemical shift perturbations arising from changes in the scFvB80 structure are indicated by the antiphase peaks observed in the difference spectra. Overall, these simple NMR experiments have confirmed that scFvB80 interacts strongly with the two linear peptides (TLGRPSGLVG and

WGFDFDFGS) previously identified by Jette et al.. Based on these preliminary results we believe that further studies are warranted using more sophisticated NMR methods and more appropriately labeled protein and/or peptides.

Homology Modeling and Docking. The above studies have shown that scFvB80 exhibits comparable activities to the full-length B80.3 antibody. Therefore, detailed structural studies of scFvB80 alone and in complex with PSA should offer an effective way to understand PSA antibody interactions at an atomic level. While detailed X-ray and NMR studies of scFvB80 and scFvB80 in complex with PSA are now underway it is likely that complete structural determination of either molecule will take quite some time and require considerable effort. Rather than await the completion of these studies we decided to make use of recently developed computer tools to generate a working hypothesis about the structure of both scFvB80 and scFvB80 as complexed with PSA. These models could serve not only as a vehicle for devising mutagenisis experiments but also to assist in the structure determination (via molecular replacement in X-ray work) or in confirming NMR assignments (through homologous assignment methods - Wishart et al., 1997). Over the past few years, homology modeling techniques have been refined to such a degree that functionally significant regions of related proteins can now be predicted with surprisingly good accuracy (Fisher et al., 1994; Bridon and Dowell, 1995; Mumenthaler et al., 1997; Lacroix et al., 1997; Sticht et al, 1997). A protein with at least 40% identity to a known structure can typically be modeled, with accuracy approaching that of a medium-resolution X-ray structure or a high-resolution NMR structure (Sanchez and Sali 1997). The availability of abundant high quality X-ray antibody structures with very high (>65%) sequence identity to scFvB80 suggests that a very good quality scFvB80 model could be generated through homology modeling. In particular, the sequence identities to the molecular templates for scFvB80 and its V_L and V_H domains were found to be 64% (two gaps), 90% (no gap) and 76% (no gap), respectively. From a PROCHECK analysis (Laskowski et al., 1994) all three models were found to exhibit good steric and geometric qualities with main-chain dihedral distributions, side-chain angle distributions, nonbonded interactions, and hydrogen bonding energies typical of a 2.4 Å resolution X-ray structure. The scFvB80 model is shown in Figure 6.11.

From this model of scFvB80 we further studied its interaction with PSA via computer-aided docking techniques. It has recently been shown that some of the better quality protein-docking algorithms can predict the structure of protein complexes with sufficient detail and accuracy that they closely match the observed or experimentally determined structures (Shoichet and Duntz 1996). Using the DockVision protocols described previously we found three sets of viable, low-energy scFvB80/PSA modes which we designated as B1, B2 and B3. It is interesting to note that in all three binding modes, at least two CDR loops (Chothia and Lesk, 1987; Chothia et al., 1989) from each V_L and V_H chain were found to interact with PSA. In the B1 cluster, CDR H2 (Ser56-Gly58) and CDR L1 (Tyr167-Asn169) bind with PSA's variable loop4 (Arg85-Leu87) and loop6 (Asp 156-Cys158) respectively. In the B2 cluster, CDR H2 (Ser56-Gly58) and CDR L1 (Tyr167-Asn169) bind with PSA residues Met116-Leu118 and Glu8-His10 respectively. In the B3 cluster, CDR H3 (Ser101- Thr104) and CDR L3 (Asn231- Asp233) bind with PSA's loop3 (Phe59-Phe61) and loop1 (Arg21-Arg23) respectively. To identify the most stable binding mode, atomic-specific hydrophobic energies (Eiserberg and McLachlan, 1986) and hydrogen bond energies were also determined. The calculated hydrophobic energies for the B1 and B3 binding modes were found to be quite close albeit somewhat lower than that of the B2 binding mode. Interestingly, the inclusion of hydrogen bond interactions actually makes the total binding energy of B3 much more favorable than either B1 or B2. These three binding modes together with their calculated "free energies of binding" are shown in Figure 6.12. To further distinguish which binding mode was most reasonable we included additional experimental information from the studies of Jette et al. (1996) and others. In particular, visual inspection of the B1 binding complex suggests that if scFvB80 were to bind at this position, it would block PSA's catalytic triad (His41, As96 and Ser71). Studies by Villoutreix et al. (1996) suggest that this site is where antichymotrypsin (ACT) typically binds to PSA. Given that earlier experiments have shown that the B80.3 Mab can bind to PSA/ACT complexes (Jette et al., 1996) these data suggest that the B1 binding mode is not a reasonable candidate. Other experimental data collected by Jette et al. (1996) using epitope scanning and phage display suggest that B80.3 probably binds to PSA's third loop (residues 50-59). Given that the B3 binding mode has the most favorable binding energy and given that it is the only model consistent with the PSA peptide scanning results (Jette et al., 1996; Jette, 1997) we believe the B3 binding pattern is experimentally and theoretically the most reasonable. Our working structural model for the scFvB80/PSA complex is, therefore, presented in Figure 6.13.

6.4. Conclusion

To summarize, the production and purification of an anti-PSA single chain antibody (scFvB80) using *P. pastoris* as an expression vehicle is reported. Competitive ELISA's indicate that this antibody fragment has approximately 20% of the PSA binding activity relative to the full-length B80.3 Mab. The protein appears to be sufficiently stable and soluble to collect good quality NMR spectra and to perform NMR-based peptide binding experiments. Our NMR data indicates that the scFvB80 antibody fragment interacts with linear peptides corresponding to PSA epitopes and mimetopes previously identified for the whole antibody. The successful development and implementation of a novel isotopic labeling strategy, described herein, suggests that sufficient quantities of isotopically labeled protein can be prepared to initiate more detailed structural studies via multidimensional heteronuclear NMR. In combination with these experimental data, computer-aided homology modeling and protein docking techniques have allowed us to prepare atomic resolution models of both the scFvB80 and the PSA-scFvB80 complex. In our model of the PSA-scFvB80 complex, residues Asn231-Asp233 and Gly102-Thr104 of scFvB80 interact with residues Ser20-Gly22 and His58-Asp61 of PSA respectively. Future structural studies (X-ray and NMR) are planned to attempt to verify these models.

References

Alfthan, K., Takkinen, K., Sizmann, D., Soderlund, H. and Teeri, T. T. (1995) *Prot. Eng.* 8, 725-731.

Bax, A. and Davis, D. G. (1985a) *J. Magn. Reson.* 65, 355-360.

Bax, A. and Davis, D. G. (1985b) *J. Magn. Reson.* 63, 207-213.

Benedict, C. A., MacKrell, A. J. and Anderson, W. F. (1997) *J. Immun. Meth.* 201, 223-231.

Brinkmann, U., Carlo, A. D., Vasmatzis, G., Kurochkina, N., Beers, R., Lee, B. and Pastan, I. (1997) *J. Mol. Biol.* 268, 107-117.

Brunger, A. T. (1992) X-PLOR 3.1 Manual, Yale University Press, New Haven, CT.

Brawer, M.K. (1991) *Acta Oncol.* 30, 161-168.

Bridon, D. P. and Dowell, B. L. (1995) *Urology* 45, 801-806.

Bruyns, A. M., De Jaeger, G., De Neve, M., De Wilde, C., Van Montagu, M. and Depicker, A. (1996) *FEBS Lett.* 386, 5-10.

Chothia, C. and Lesk, A. M. (1987) *J. Mol. Biol.* 196, 901-917.

Chothia, C., Lesk, A. M., Tramontano, A., Levitt, M., Smith-Gill, S. J., Air, G., Sheriff, S., Paddlan, E. A., Davies, D. and Tulip, W.R. (1989) *Nature* 342, 877-883.

Campbell, A. P., Yong, W. Y., Houston, M. J., Schweizer, F., Cachia, P. J., Irvin, R. T., Hindsgaul, O., Hodges, R. S. and Sykes, B. D. (1997) *J. Mol. Biol.* 267, 382-402.

Devlin, C. M., Bowles, M. R., Gordon, R. B. and Pond, S. M. (1995) *J. Biochem.* 118, 480-487.

Eisenberg, D and McLachlan, A. D. (1986) *Nature* 319, 199-203.

Eldin, P., Pauza, M. E., Hieda, Y., Lin, G., Murtaugh, M. P., Pentel, P. R. and Pennell, C. A. (1997) J. *Immu. Meth.* 201, 67-75.

Evan, G. J., Lewis, G. K., Ramsay, G. and Bishop, J. M. (1985) *Mol. Cell. Biol.* 5, 3610-3616.

Fisher, C. L., Greengard, J. S. and Griffin, J. H. (1994) *Science* 264, 82-85.

Freund, C., Ross, A., Pluckthun, A. and Holak, T. A. (1994) *Biochemistry* 33, 3296-3303.

Freund, C., Ross, A., Guth, B, Pluckthun, A. and Holak, T. A. (1993) *FEBS Lett.* 320, 97-100.

Graves, H. C. B. (1993) *Clin. Chem.* 37, 1618-1625.

Hart, T. N., Ness, S. R. and Read, R. J. (1997) *Proteins: Structure, Function, and Genetics, Suppl.* 1, 205-209.

Hart, T. N. and Read, R. J. (1992) *Proteins: Structure, Function, and Genetics* 13, 206-222.

Huang, X., Yang, X., Luft, B. J. and Koide, S. (1998) *J. Mol. Biol.* 281, 61-67.

Hubbard, J. A. M., Raleigh, D. P., Bonnerjea, J. R. and Dobson, C. M. (1997) *Prot. Sci.* 6, 1945-1952.

Irani, J., Millet, C., Levillain, P., Dore, B., Begon, F. and Aubert, J. (1996) *Eur. Urology* 29, 407-412.

Jeener, J., Meier, B. H., Bachmann, P. and Ernst, R. R. (1979) *J. Chem. Phys.* 71, 4546-4553.

Jette, D. C., Kreutz, F. T., Malcolm, B. A., Wishart, D. S., Noujain, A. A. and Suresh, M. R. (1996) *Clin. Chem.* 42, 1961-1969.

Jette, D. C. (1997) Ph. D Thesis, University of Alberta, Edmonton, Alberta, Canada.

Kay, L. E., Keifer, P. and Saarien, T. (1992) *J. Am. Chem. Soc.* 114, 10663-10665.

Kipriyanov, S. M., Moldenhauer, G. and Little, M. (1997) *J. Imm. Meth.* 200, 69-77.

Kortt, A. A., Malby, R. L., Caldwell, J. B., Gruen, C., Ivancic, N., Lawrence, M. C., Howlett, G. J., Webster, R. G., Hudson, P. J. and Colman, P. M. (1994) *Eur. J. Biochem.* 221, 151-157.

Kreutz, F. T. and Suresh, M. R. (1995) Fourth international conference on bispecific monoclonal antibodies and targeted cellula toxicity proceedings. Hawk's Key, Florida, USA, P.17.

Kumar, A., Ernst, R. R. and Wuthrich, K. (1980) *Biochem. Biophys. Res. Commun.* 95, 1-6.

Lacroix, M., Rossi, V., Gaboriaud, C., Chevallier, S., Jaquinod, M., Thielens, N. M., Gagnon, J. and Arlaud, G. J. (1997) *Biochemistry* 36, 6270-82,

Laroche, Y., Strome, V., DeMeuter, J., Messens, J. and Lauwereys, M. (1994) *Bio/Technology* 12, 1119-1124.

Laskowski, R. A., Rullmannn, J. A., MacArthur, M. W. Kaptein, R. and Thornton, J. M. (1996) *J. Biomol. NMR* 8, 477-486.

Lengauer, T. and Rarey, M. (1996) *Current Opinion in Structural Biology* 6, 402-406.

Lilja, H., Christensson, A., Dahlen, U., Matikainen, M-T., Nilsson, O., Pettersson, K. and Lovgren, T. (1991) *Clin. Chem.* 37, 1618-1625.

Loewen, M. C., Liu, X., Davies, P. L. and Daugulis (1997) *Appl. Microbiol Biotechnol.* 48, 480-486.

Luderer, A. A., Chen, Y. T., Soriano, T. F., Kramp, W. J., Carlson, G., Cuny, C., Sharp, T., Smith, W., Petteway, J. and Brawer, M. K. (1995) *Urology* 46, 187-194.

Luo, D., Mah, N., Krantz, M., Wishart, D. S., Jacobs, F. and Martin, L. (1997a) *Biotechnology Techniques* 11, 759-761.

Luo, D., Mah, N., Krantz, M., Wilde, K., Wishart, D. S., Zhang, Y., Jacobs, F. A., and Martin, L. (1995) *J. Biochem.* 118, 825-831.

Luo, D., Mah, N., Wishart, D. S., Zhang, Y., Jacobs, F. A. and Martin, L. (1996) *J. Biochem.* 12, 229-232.

Luo, D., Geng, M., Noujaim, A. A. and Madiyalakan, R. (1997b) *J. Biochem.* 121, 831-834.

Madden, T. L., Tatusov, R. L. and Zhang, J. (1996) *Meth. Enzymol.* 266, 131-141.

Malby, R. L., McCoy, A. J., Kortt, A. A., Hudson, P. J . and Colman, P. M. (1998) *J. Mol. Biol.* 279, 901-910.

Marshall, G. R., Gorin, F. A. and Moore, M. L. (1978) *Annu. Rev. Med. Chem.* 13, 227-238.

Milenic, D. E., Yokota, T., Filpula, D. R., Finkelman, M. A. J., Dodd, S. W., Wood, J. F., Whitlow, M., Snoy, P. and Schlom, J. (1991) *Cancer Research* 51, 6363-6371.

Munro, S. and Pelham, H. (1986) *Cell* 46, 291-300.

Munenthaler, C., Schneider, U., Buchholz, C., Koller, D., Braun, W. and Cattaneo, R. (1997) *Prot. Sci.* 6, 588-597.

Nicholson, I. C., Lenton, K. A., Little, D. J., Decorso, T., Lee, F. T., Scott, A. M., Zola, H. and Hohmann, A. W. (1997) *Molecular Immunology* 34, 1157-1165.

Piironen, T., Villoutreix, B. O., Becker, C., Hollingsworth, K., Vihenen, M., Bridon, D., Qiu, X., Rapp, J., Dowell, B., Lovgren, T., Pettersson, K. and Lilja, H. (1998) *Prot. Sci.* 7, 259-269.

Ploch, N. R. and Brawer, M. K. (1994) *Urology* 43, 27-35.

Reiter, Y., Brinkmann, U., Jung, S. H., Lee, B. and Pastan, I. (1994a) *Prot. Eng.* 7, 697-704.

Reiter, Y., Brinkmann, U., Jung, S.H., Lee, B., Kasprzyk, P. G., King, C. R. and Pastan, I. (1994b) *J. Biol. Chem.* 269, 18327-18331.

Raag, R. and Whitlow, M. (1995) *FASEB* 9, 73-80.

Ridder, R., Schmitz, R., Legay, F. and Gram, H. (1995) *Bio/Technology* 13, 255-260.

Sanchez, R. and Sali, A (1997) *Current Opinion in Structural Biology* 7, 206-214.

Seregni, E., Botti, C., Ballabio, G. and Bombardieri E. (1996) *Tumori* 82, 72-77.

Schulze, R. A., Kontermann, R. E., Queitsh, I., Dubel, S. and Bautz, E. K. F. (1994) *Anal. Biochem.* 220, 212-214.

Shoichet, B. K. and Kuntz, I. D. (1996) *Chemistry and Biology* 3, 151-156.

Stader, J. A. and Silhavy, T. (1990) *Meth. Enymol.* 185, 166-186.

Sticht, H., Gallert, K. C., Krauss, G. and Rosch, P. (1997) *Journal of Biomolecular Structure & Dynamics* 14, 667-675.

Tsang, P., Rance, M., Fieser, T. M., Ostresh, J. M., Houghten, R. A., Lerner, R. A. and Wright, P. E. (1992) *Biochemistry* 31, 3862-3871.

Villoutreix, B. O., Lilja, H., Pettersson, K., Lovgren, T. and Teleman, O. (1996) *Prot. Sci.* 5, 836-851.

Villoutreix, B. O., Getzoff, E. D. and Griffin, J. H. (1994) *Prot. Sci.* 3, 2033-2044.

Vriend, G. (1990) *Journal of Molecular Grahpics* 8, 52-56.

Wishart, D. S., Bigam, D. G., Holm, A., hodges, R. S. and Sykes, B. D. (1995) *J. Biolmol. NMR* 5, 67-81.

Wishart, D. S., Willard, L. and Sykes, B. D. (1995) VADAR 1.1 - University of Alberta (ftp://redpoll.pharmacy.ualberta.ca)

Wuthrich, K. (1986) *NMR of Proteins and Nucleic Acids,* Wiley, New York.

Wu, J. T. (1994*) J. Clin. Lab. Analysis* 8, 52-62.

Zdanov, A., Li, Y, Bundle, D. R., Deng, S. J., MacKenzie, R., Narang, S. A., Young, N. M. and Cygler, M. (1994). *Proc. Natl. Acad. Sci. USA.* 91, 6423-6427.

Zhang, Y., Kreutz, F. T., Luo, D., Jacobs, F. A., Martin, L., Suresh, M. R. and Wishart, D. S. (1997) in Miami *Bio/technology Short Reports* 8, p9.

Zhang, Y., Wang, Y., Kreutz, F. T., Luo, D., Jacobs, F. A., Suresh, M. R. and Wishart, D. S. (submitted)

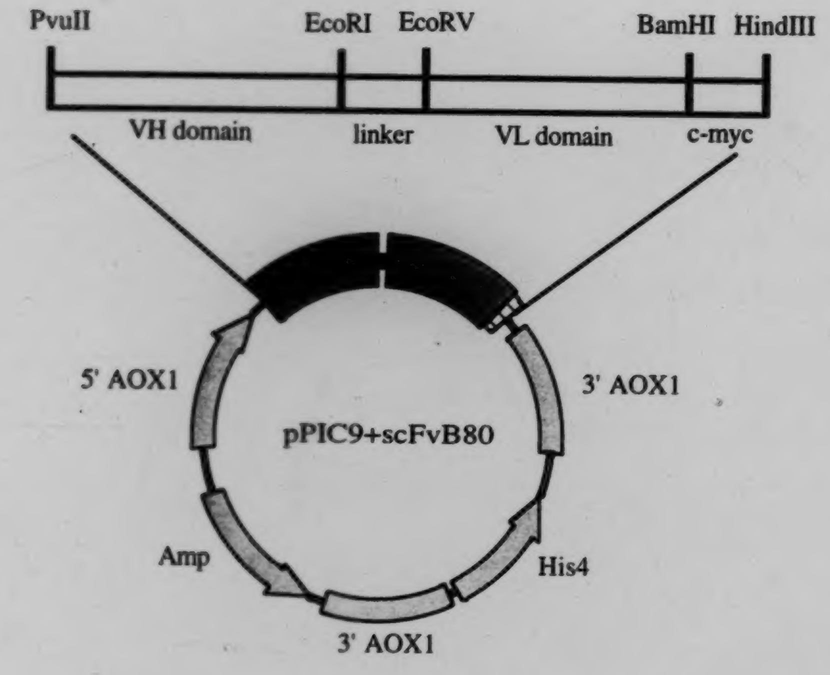

PvuII
EcoRI
EcoRV
BamHI
HindIII
VH domain
linker
VL domain
c-myc
5' AOX1
3' AOX1
pPIC9+scFvB80
Amp
His4
3' AOX1

Figure 6.1. pPIC 9 expression vector for *P. pastoris* showing the orientation and placement of the scFvB80 gene.

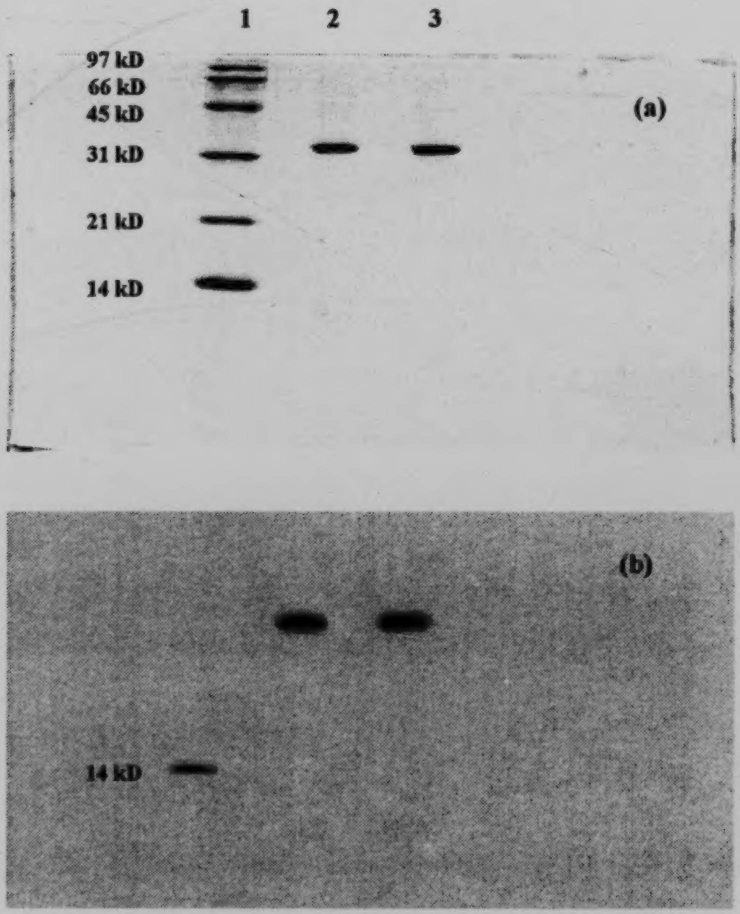

1
2
3
97 kD
66 kD
45 kD
31 kD
21 kD
14 kD
(a)
(b)
14 kD

Figure 6.2. (a) 15% SDS-PAGE gel and (b) Western Blotting of scFvB80 expressed in *P. pastoris*. Lane 1: molecular weight marker; lane 2: culture supernatant; lane 3: purified scFvB80.

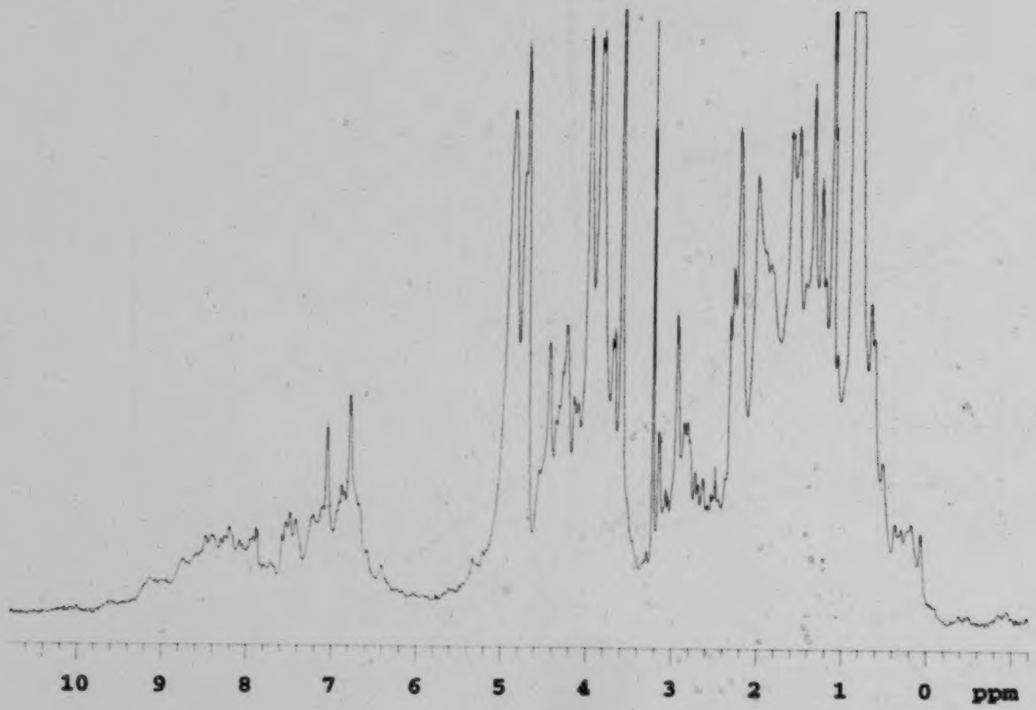


Figure 6.3. One-dimensional ^{1}H NMR spectrum of scFvB80 showing the high purity of the sample. The downfield α-proton peaks (5-6 ppm) and the upfield side chain proton peaks (0 to -2 ppm) indicate that the protein exhibits a well-folded structure.

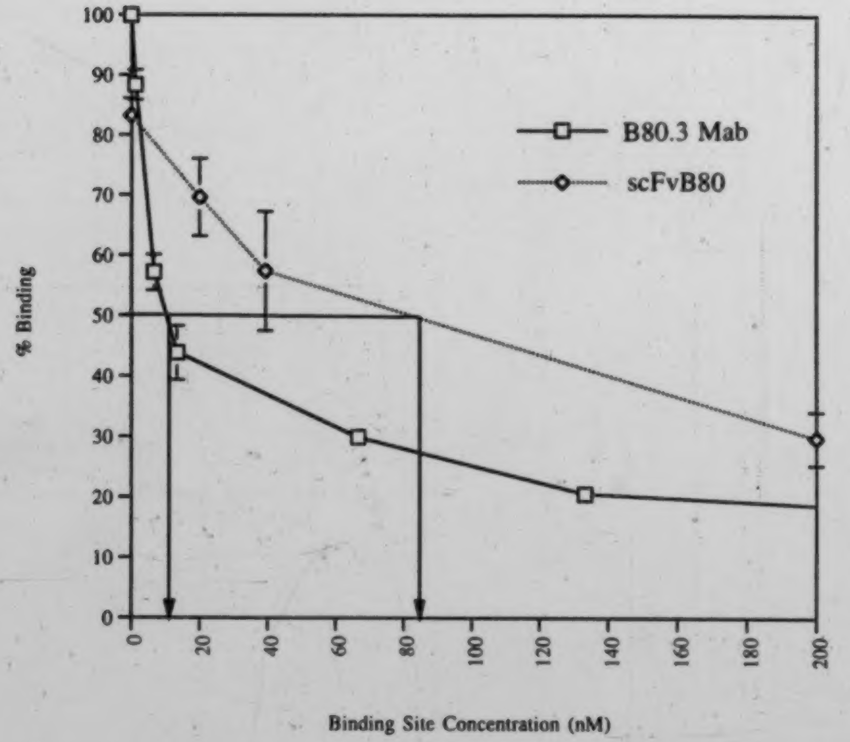


Figure 6.4. Competitive ELISA of B80.3 Mab and scFvB80. The IC50 of B80.3 Mab was found to be approximately 15 nM and that for scFvB80 approximately 85 nM by this method.

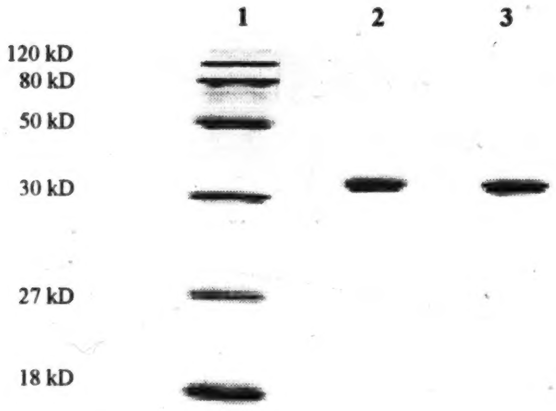

1
2
3
120 kD
80 kD
50 kD
30 kD
27 kD
18 kD

Figure 6.5. 15% SDS-PAGE gel illustrating the improved level of ^{15}N labeled protein production using our specially developed labeling media for *P. pastoris*. Lane1: molecular weight marker; lane 2: BMGY media; lane 3: home made media.

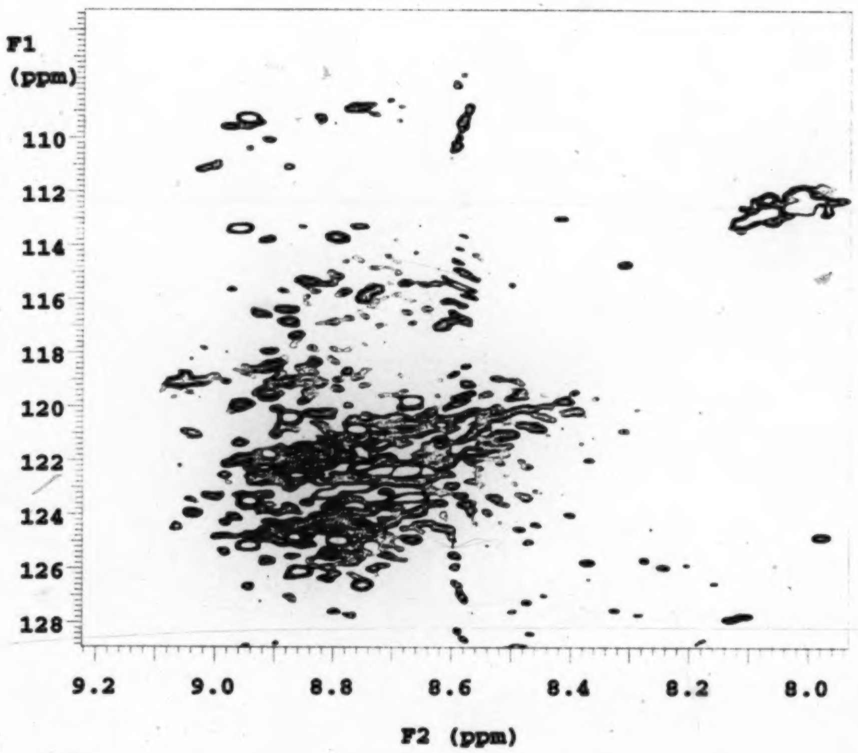


Figure 6.6. Selected region of ^{15}N HSQC NMR spectrum of ^{15}N labeled scFvB80.

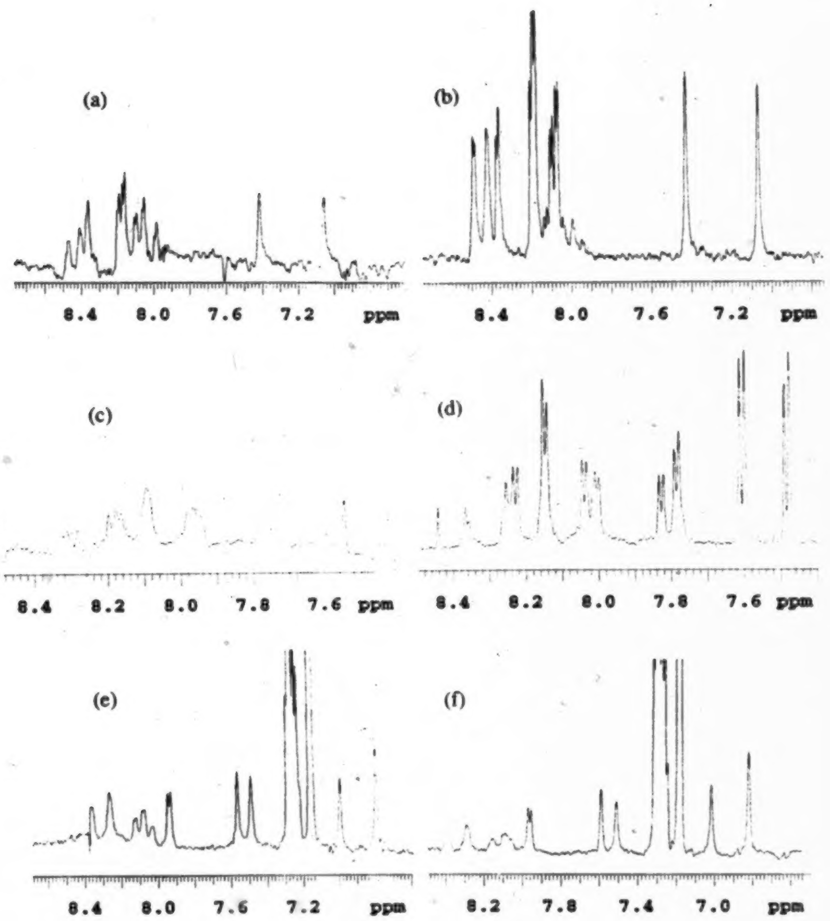


Figure 6.7. ^{1}H spectra showing the resonance intensity changes occurring when various peptides are added to a solution containing scFvB80. Shown in this figure are spectra corresponding to: (a) the corresponding difference spectra [(scFvB80 + TLGRPSGLVG) - scFvB80]; (b) the PSA epitope peptide (TLGRPSGLVG) alone; (c) the corresponding difference spectra (scFvB80 + WGFDFGFGS) - scFvB80; (d) the PSA mimetope peptide (WGFDFGFGS) alone; (e) the corresponding difference spectra [(scFvB80 + SRLKNFVRE) - scFvB80]; (f) the non-binding control peptide (SRLKNFVRE) alone.

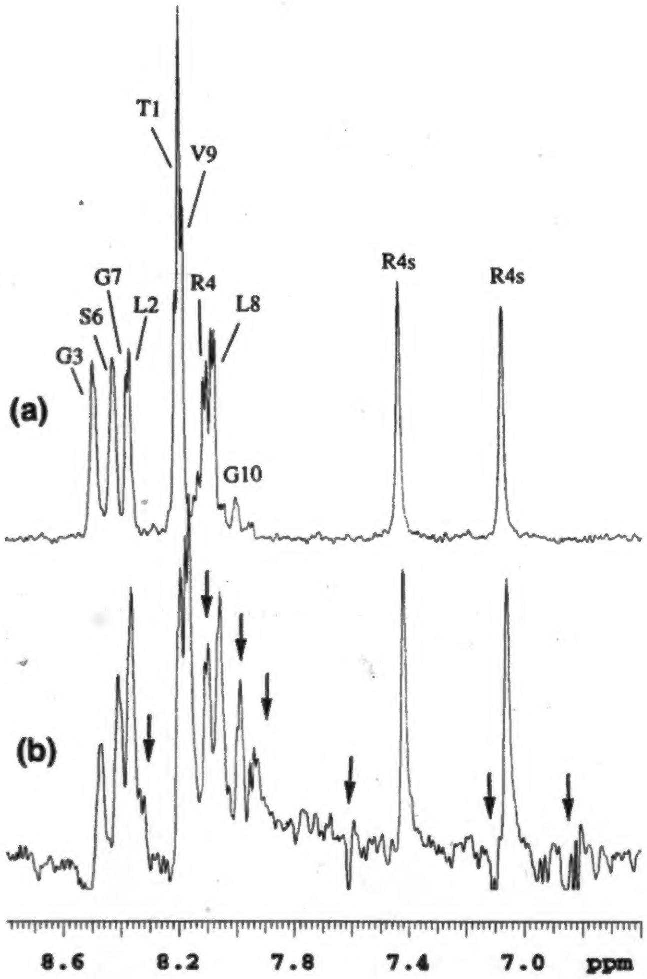


Figure 6.8. ^{1}H NMR spectra of (a) the PSA epitope analog peptide (TLGRPSGLVG) free in solution and (b) the result of subtracting the ^{1}H spectrum of free scFvB80 from the ^{1}H spectrum of the scFvB80-peptide complex. The major differences are highlighted by arrows.

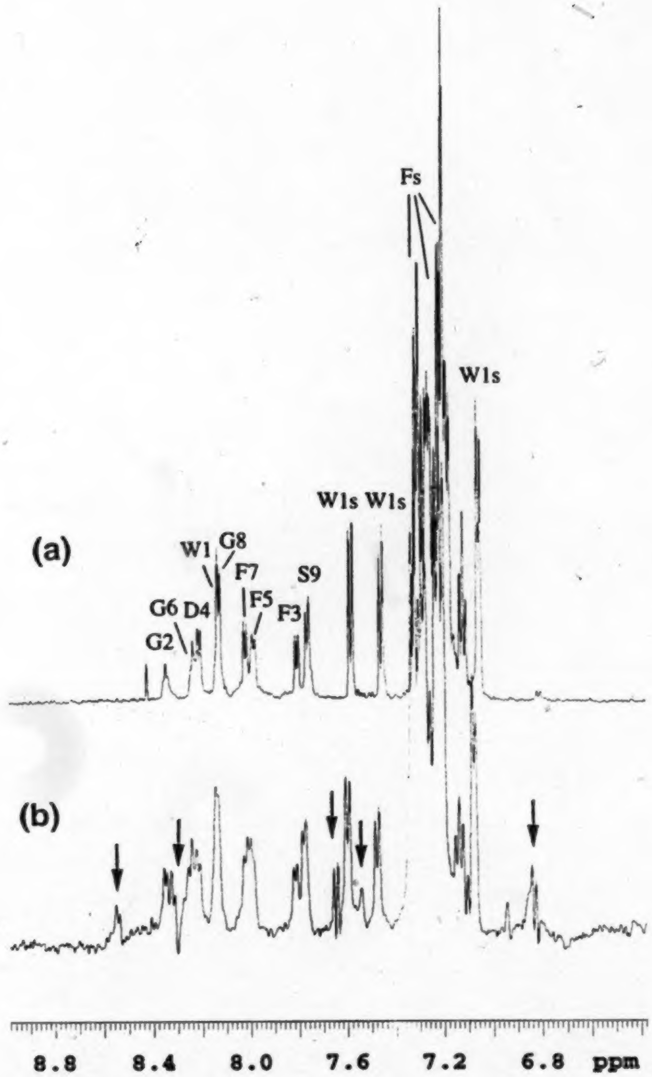


Figure 6.9. ^{1}H NMR spectra of (a) the PSA mimetope (WGFDFGFGS) free in solution and (b) the result of subtracting the ^{1}H spectrum of free scFvB80 from the ^{1}H spectrum of scFvB80-peptide complex. The major differences are highlighted by arrows.

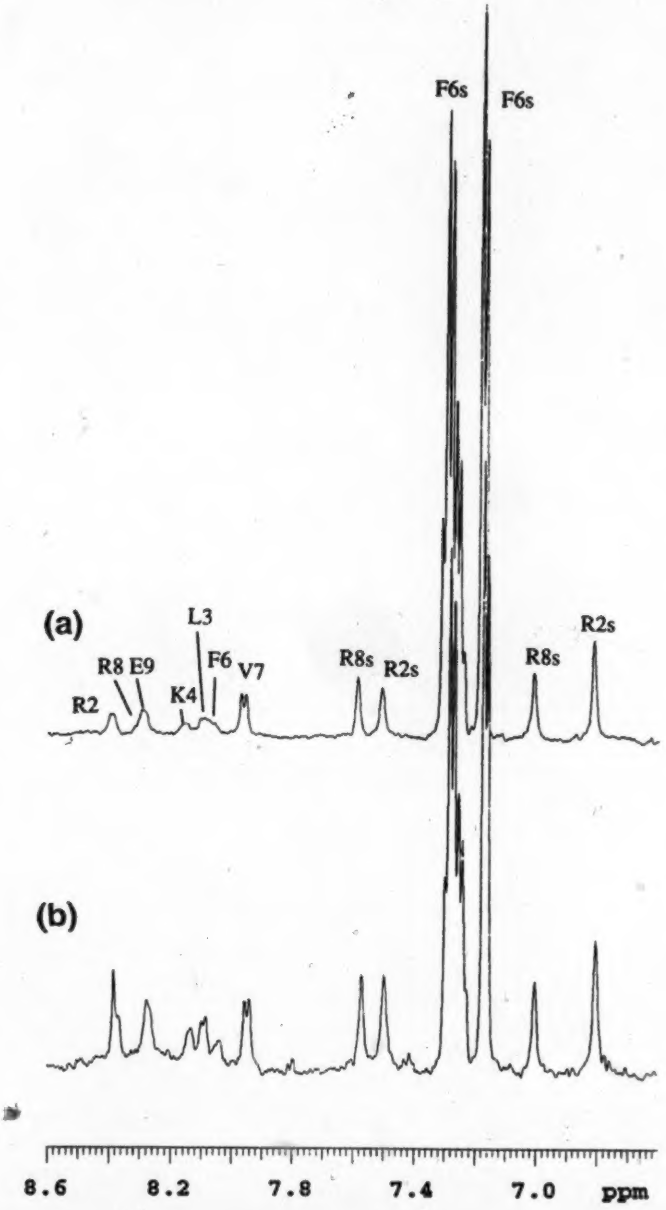


Figure 6.10. ^{1}H NMR spectra of (a) control peptide (SRLKNFVRE) free in solution and (b) the result of subtracting the ^{1}H spectrum of free scFvB80 from the ^{1}H spectrum of scFvB80-peptide complex.

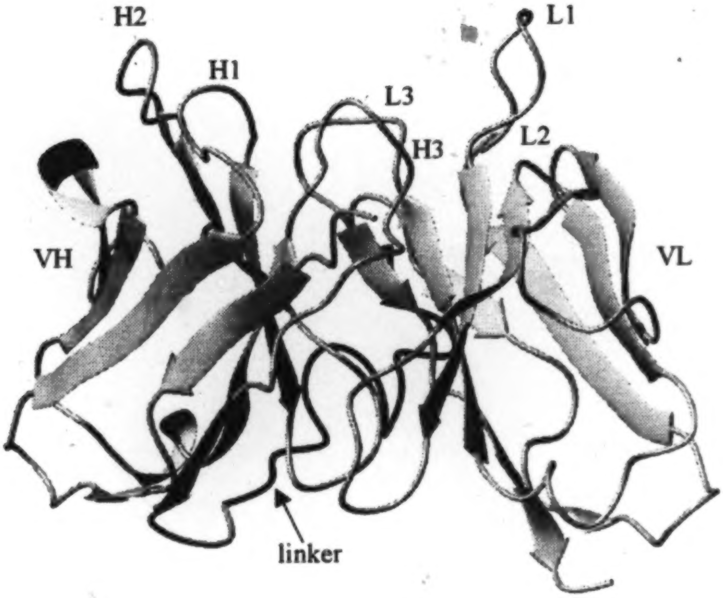

H2
H1
L3
H3
L1
L2
VH
VL
linker

Figure 6.11. Ribbon diagram of the 3D model of scFvB80. The linker is marked in black

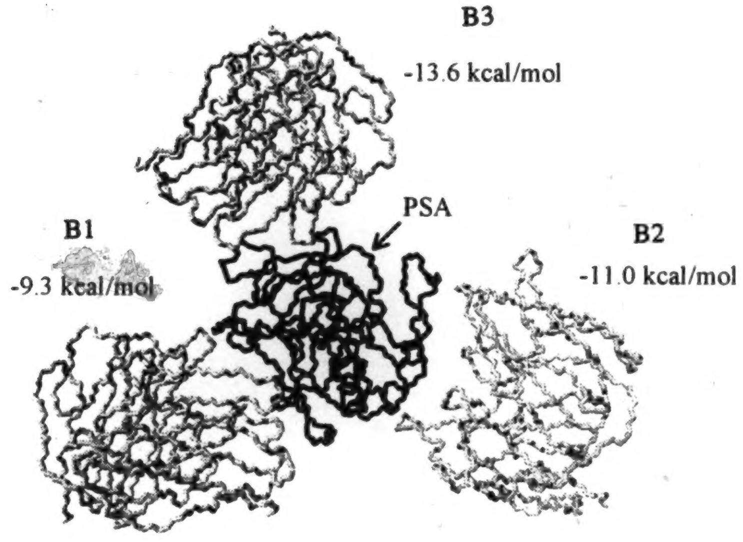

B3
-13.6 kcal/mol
PSA
B1
-9.3 kcal/mol
B2
-11.0 kcal/mol

Figure 6.12. Three energetically favorable scFvB80/PSA binding clusters from 1000 independent Monte Carlo simulation searches by DockVision. The calculated free energies of binding for each each cluster are indicated.

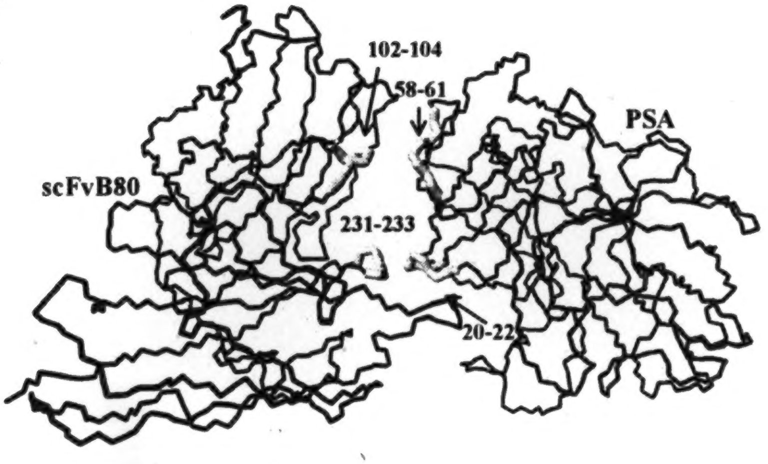

102-104
58-61
PSA
scFvB80
231-233
20-22

Figure 6.13. The 3D model of the scFvB80/PSA complex. In this model, the CDR H3 (102-104) and L3 (231-233) of scFvB80 bind with the variable loop3 (58-61) and loop1 (20-22) of PSA repectively.

Chapter 7

General Discussion and Conclusion

The intent of this thesis was to describe several novel methods that could be used to assist in the physical characterization of pharmaceutically important proteins. Specifically, in Chapters 2 and 3, I described two new methods for the quantitative measurement of $^3J_{HNH\alpha}$ coupling constants and in Chapter 6 I described a new and cost-effective method for the isotopic labeling of proteins expressed in *P. pastoris*. These methods were applied to the NMR characterization of three proteins of potential pharmaceutical importance -- T4 Grx, CbnB2 and scFvB80 -- and the results are summarized in Chapters 4, 5 and 6 along with detailed data concerning the solution structure of these proteins. It is my contention that these methods, in conjunction with other NMR spectroscopic techniques used to generate their solution structures have resulted in an improved understanding of the structure/function relationships for these three interesting and potentially useful proteins.

As I described previously in Chapters 2 and 3, coupling constant measurements can play an important role in the determination and refinement of peptide and protein structures (Bax et al., 1994). Specifically, the magnitude of three bond HN and H_α coupling constant, $^3J_{HNH\alpha}$, can be directly related to the polypeptide dihedral angle phi through the well-known Karplus equation (Wang and Bax, 1996). Over the past two decades, a variety of techniques have been developed to quantitatively measure $^3J_{HNH\alpha}$ from both homonuclear and heteronuclear NMR spectra. The principles of these techniques can be classified into the following: (1) measurement of peak to peak separation (Pardi et al., 1984; Kim et al., 1989; Kay et al., 1989; Smith et al., 1991); (2) computer-aided curve fitting (Szyperski et al., 1992; Ludvigsen et al., 1991; Kay and Bax, 1990; Goodgame and Greer, 1993); and (3) peak volume integration (Billeter et al., 1992; Vuister and Bax, 1993; Weismann et al., 1994). Quantitative coupling constant measurements which use these conventional methods often require special hardware or special computer programs, the implementation of a complex pulse sequence, the use of isotopically labeled material, or the collecting and reprocessing of multiple data sets. These complicated requirements often make it difficult to obtain the quantitative value for this important parameter.

This thesis reports a simple linear relationship between the $^3J_{HNH\alpha}$ coupling constant and the linewidth ($\Delta\nu_{1/2}$) of in-phase NMR peaks. This observation led to the development of two novel methods for rapid and accurate determination of polypeptide $^3J_{HNH\alpha}$ coupling constants from commonly used and easily obtainable TOCSY, NOESY and (^{15}N) HMQC-J spectra. Chapter 2 of this thesis describes how to extract $^3J_{HNH\alpha}$ coupling constants from simple inspection of amide cross peaks in homonuclear ^1H TOCSY or ^1H NOESY spectra. By using the appropriate set of processing parameters it was shown that $^3J_{HNH\alpha} = 0.5\ (\Delta\nu_{1/2}) - MW/5000 + 1.8$ for TOCSY spectra and $^3J_{HNH\alpha} = 0.6\ (\Delta\nu_{1/2}) - MW/5000 - 0.9$ for NOESY spectra, where $\Delta\nu_{1/2}$ is the half-height linewidth in Hz and MW is the molecular weight of the protein in Da. The simplicity of this relationship, combined with the ease with which $\Delta\nu_{1/2}$ measurements can be made allows $^3J_{HNH\alpha}$ coupling constants to be quickly determined without the need for any complex curve-fitting algorithms. Tests on 11 different polypeptides involving more than 650 separate $^3J_{HNH\alpha}$ measurements have shown that this method yields coupling constants with an rmsd error (relative to X-ray data) of less than 0.9 Hz. The correlation coefficient between the predicted NMR coupling constants and those derived from high-resolution X-ray crystal structures is typically better than 0.89. These simple linear relationships have been found to be valid for peptides as small as 1 kDa and proteins as large as 20 kDa. Despite the method's simplicity, these results are comparable in accuracy and precision to the best techniques published to date. In Chapter 3, this linear relationship is further extended to extract the $^3J_{HNH\alpha}$ coupling constants of ^{15}N labeled proteins from HMQC-J spectra. The HMQC-J experiment is one of the most popular methods for extracting $^3J_{HNH\alpha}$ coupling constants (Kay et al., 1989; Kay and Bax, 1990). By first measuring how the peak-to-peak separation changes as a function of the line-narrowing filters and then fitting these results to a simulated cure, it is possible to determine $^3J_{HNH\alpha}$ values with relatively good accuracy. However, this iterative fitting process is both tedious and error-prone. Using appropriately processed HMQC-J data, it was demonstrated that a simple linear relationship exists between the half-height linewidth ($\Delta\nu_{1/2}$) of ^{15}N-^1H cross peaks and their corresponding $^3J_{HNH\alpha}$ coupling constants. Tests on more than 800 data points from six different protein HMQC-J spectra indicate that this technique permits accurate measurements of up to 100 $^3J_{HNH\alpha}$ coupling constants in less than 30 min. Comparisons between $^3J_{HNH\alpha}$ values predicted from high resolution X-ray structures and those determined using this technique indicate that the method is both accurate and precise (correlation coefficient = 0.90, rmsd = 0.75 Hz).

This new concept of linewidth measurement makes quantitative coupling constant measurements far simpler and far easier to use in analyzing the solution conformation of peptides and proteins via NMR. For future work, it will be of interest to see if this linear relationship can be used to measure other coupling constants (e.g., αβ coupling constants) of peptides, proteins, polynucleotides or other biomolecules. It may also be of great practical importance to test whether this linear relationship between the half-height linewidth and coupling constants still exists in three- and four-dimensional NMR spectra. As described in Chapters 4 and 5, this linewidth measurement greatly facilitated the structural determination of T4 Grx and CbnB2.

Chapter 4 reports on the solution structures of both the reduced and oxidized forms of wild-type bacteriophage T4 glutaredoxin as determined by conventional two-dimensional NMR spectroscopy. A total of thirty structures were generated for each redox state by means of a simulated annealing protocol, using more than 1500 experimental NMR restraints. As expected, the overall fold of both oxidized and reduced T4 Grx was very similar to the previously determined 1.45 A resolution X-ray structure of oxidized T4 Grx (PDB: 1ABA). Both redox forms of T4 Grx are characterized by a central four-stranded β-sheet surrounded by three helices and two flexible loops. Upon reduction, several notable changes were observed, including (1) chemical shift changes for several residues proximal to the redox-active site; (2) increased spatial separation between the two cysteines; and (3) a marked narrowing of resonance linewidth, indicating a dimer-to-monomer transition. The overall root mean square deviation (RMSD) value relative to the mean coordinates for the final set of 30 oxidized structures was 0.71 ± 0.1 Å (backbone) and 1.27 ± 0.11Å (heavy atoms). For the final set of 30 reduced structures, the RMSD was 0.70 ± 0.11 Å (backbone) and 1.28 ± 0.11 Å (heavy atoms).

In Chapter 4, I also attempt to address the question of whether T4 Grx is really a glutaredoxin or a thioredoxin. Unlike any other glutaredoxin or thioredoxin, T4 Grx seems to share a dual role as it can be reduced not only by glutathione but also by thioredoxin reductase (Holmgren, 1989). Due to this hybrid functional property, this protein was formerly known as a bacteriophage T4 thioredoxin. However, sequence comparisons suggest that it should belong to the glutaredoxin family (Nikkola et al., 1991). On the other hand, measurement of the thiol ionization constants of reduced T4 Grx by ultraviolet absorbance spectroscopy indicate that the pKa values of Cys 14 and Cys 17 are 6.8 and 9.5 respectively. The unusually high pKa value of Cys14 suggests that T4 Grx may actually share more functional similarity with the thioredoxin superfamily than with the glutaredoxin superfamily. The fact that bacteriophage T4 Grx exhibits features of both thioredoxin and glutaredoxin but lacks sequence similarity to any other glutaredoxin or thioredoxin indicates that this protein essentially falls between the two major classes. This suggests that the evolutionary placement of this unique protein will likely require further structural and functional studies. I believe that the study presented in Chapter 4 has helped to lay a foundation for future structural and dynamic studies of this unique and potentially protein.

Chapter 5 reports on the solution structures of Carnobacteriocin B2 in 90% TFE. CbnB2 is a 48 residue type IIa bacteriocin (i.e. an antimicrobial peptide) isolated from the lactic acid bacterium (*Carnobacterium pisicola*). Type IIa bacteriocins are normally characterized by the presence of a well-conserved heptapeptide (YGNGVXC) sequence motif and an equally well-conserved disulfide bond located near the N-terminus of the peptide. Carnobacteriocin B2 appears to be unique among this class of bacteriocins in that its two conserved cysteines (Cys9 and Cys14) apparently do not form a disulfide bond (Quadri et al., 1997). In an effort to determine the structural consequences of this non-existent disulfide and to further determine if the YGNGVXC sequence motif confers a conserved "receptor-binding" structural motif (Fleury et al., 1996), the three-dimensional solution structure of Carnobacteriocin B2 was determined using standard two-dimensional ^{1}H NMR techniques. Additional mass spectroscopic and thiol modification experiments, in conjunction with detailed activity measurements, were conducted to verify the redox status of CbnB2. These later experiments indicate that CbnB2, in contrast to earlier reports, does indeed possess a disulfide bond and this disulfide bond is essential to antimicrobial activity. Furthermore, our NMR results indicate that CbnB2, in 90% trifluoroethanol, has a well-defined central helix (from residues 18 to 39), but a disordered N terminus.

Comparisons of the CbnB2 structure with the refined solution structure of Leucocin A – another type IIa bacteriocin—indicate that this helical structure is well conserved between the two peptides, but the N-terminal structure (i.e. the presumptive receptor-binding site) is not. This result is of particular interest because LeuA and CbnB2 actually exhibit >66% sequence identity over their first 24 residues. This suggests that the highly ordered N-terminal β-sheet structure seen in LeuA may either be an artifact of the solution conditions (90% TFE) or that the N-terminal structure (i.e. the presumptive receptor binding site) of type IIa bacteriocins may not be critical to their function.

Chapter 6 reports the high level production of an anti-PSA single chain antibody (scFv B80) using *Pichia pastoris* as the expression vehicle. SDS-PAGE indicates that this recombinant antibody is the only protein detectable in the growth media. The production (>15 mg/L) and purity of the secreted and correctly folded protein is sufficiently high to permit a simple three-step purification process. Antibody-antigen affinity tests using competitive ELISA indicate that this antibody fragment has approximately 20% of the activity of the full-length antibody. The interaction between scFvB80 and previously identified PSA epitope/mimetope peptides (Jette, 1997) was also studied by one-dimensional ^{1}H NMR. The difference spectra between the proton NMR spectra of the peptide-scFvB80 complex and that of the scFvB80 confirm that scFvB80 interacts, albeit weakly, with these linear peptides. To further characterize the strength of the interaction between scFvB80 and the PSA epitope/mimetope peptides, isotope-edited NMR experiments (Tsang et al., 1992; Campbell et al., 1997) are suggested for future work.

The above experimental data, along with preliminary models of scFvB80 (both free and complexed with PSA) have allowed us to develop a working hypothesis regarding the interaction of PSA with its targeting antibodies. Our high quality structural model of scFvB80, along with its two individual V_H and V_L domains should be of help in completing further structural studies by NMR. The model for the PSA/scFvB80 complex will also be a valuable test of the predictive abilities of DockVison. Chapter 6 also describes a new and potentially useful method for preparing isotopically labeled protein in *P. pastoris*. Previous efforts describing the production of isotopically labeled protein in *P. pastoris* have either resulted in a substantial reduction in yields (>10 fold) using shaker flasks and a defined minimal media (Laroche et al., 1994) or have only been able to succeed in expressing comparable yields using specially optimized large volume (i.e. expensive) fermentation systems (Loewen et al., 1997; Wood and Komives, 1999) with defined minimal media. Chapter 6 reports a new protocol to produce isotopically labeled protein from *P. pastoris* in which an easily prepared 'home-made' ^{15}N labeled yeast extract serves as a rich-medium substitute for minimal or defined media. This particular protocol has two key advantages: first, it is amenable to small-scale, inexpensive shaker flask growths and second, it allows one to continuously recycle expensive isotopic materials. The results to date are promising, with good yields demonstrated for three different proteins expressed in *P. pastoris* (two different antibody fragments and an antifreeze protein). However, it is likely that further optimization is needed to make it a "universal" approach for isotopic labeling. It would also be of interest to see if this "home-made" media could also be used in a fermentor-based system and if protein yields could be substantially improved over what has been obtained in shaker flasks to date.

References

Bax, A., Vuister, G. W., Grzesiek, S., Delaglio, F., Wang, A. C., Tschudin, R. and Zhu, G. (1994) *Meth. Enzymol.* 239, 79-105.

Billeter, N., Neri, D., Otting, G., Qian, Y. Q. and Wuthrich, K. (1992) *J. Biomol. NMR* 2, 257-274.

Campbell, A. P., Wong, W. Y., Houston Jr. M., Schweizer, F., Cachia, P. J., Irvin, R. T., Hindsgaul, O., Hodges, R. S. and Sykes, B. D. (1997) *J. Mol. Biol.* 267, 382-402.

Cregg, J. M., Vedvick, T. S. and Raschke, W. C. (1993) *Bio/Technology* 11, 905-910.

Fleury, Y., Dayem, M. A., Montagne, J. J., Chaboisseau, E., LeCaer, J. P., Nicolas, P. and Delfour, A. (1996) *J. Biol. Chem.* 271, 14421-14429.

Gardner, K. H. and Kay, L. E. (1998) *Annu. Rev. Biophys. Biomol. Struct.* 27, 357-406.

Goodgame, M. M. and Greer, S. M. (1993) *J. Magn. Reson. A.* 102, 246-248.

Holmgren, A. (1989) *J. Biol. Chem.* 264, 13963-13966.

Jette, D. C. (1997) Ph. D. Thesis, University of Alberta.

Kay, L. E. and Bax, A. (1990) *J. Magn. Reson.* 86, 110-126.

Kay, L. E., Brooks, B., Sparks, S. W., Torchia, D. A. and Bax, A. (1989) *J. Am. Chem. Soc.* 111, 5488-5490.

Kim, Y. and Prestegard, J. H. (1989) *J. Magn. Reson.* 84, 9-13.

Laroche, Y., Strome, V., DeMeuter, J., Messen, J. and Lauereys, M. (1994) *Bio/Technology* 12, 1119-1124.

Loewen, M. C., Liu, X., Davies, P. L. and Daugulis, A. J. (1997) *Appl. Microbiol. Biotechnol.* 48, 480-486.

Ludvigsen, S., Andersen, K. V. and Poulsen, F. M. (1991) *J. Mol. Biol.* 217, 731-736.

Nikkola, M., Gleason, F., Saarinen, M., Joelson, T., Bjornber, O. and Eklund, H. (1991) *J. Biol. Chem.* 266, 16105-16112.

Pardi, A., Billeter, M. and Wuthrich, K. (1984) *J. Mol. Biol.* 180, 741-751.

Quadri, L. E. N., Yan, L. Z., Stiles, M. E. and Vederas, J. C. (1997) *J. Biol. Chem.* 272, 3384-3388.

Smith, L. J., Sutcliffe, M. J., Redfield, S. C. and Dobson, C. M. (1991) *Biochemistry* 30, 986-996.

Szyperski, T., Guntert, P., Otting, G. and Wuthrich, K. (1992) *J. Magn. Reson.* 99, 552-560.

Tsang, P., Rance, M., Fieser, T. M., Ostresh, J. M., Houghten, R. A., Lerner, R. A. and Wright, P. E. (1992) *Biochemistry* 31, 3862-3871.

Vuister, G. W. and Bax, A. (1993) *J. Am. Chem. Soc.* 113, 7772-7777.

Wang, A.C. and Bax, A. (1996) *J. Am. Chem. Soc.* 118, 2483-2494.

Weisemann, R., Ruterjans, J., Schalbe, H., Schleucher, J., Bermel, W. and Griesinger, C. (1994) *J. Biomol. NMR* 4, 231-240.

Wood, M. J. and Komives, E. A. (1999) *J. Biomol. NMR* 13, 149-159.